

Edited by **MATTHEW EVAN DAVIS,**
TAMSYN MAHONEY-STEEL,
and **ECE TURNATOR**

ARC HUMANITIES PRESS

VISIT...

LANZAROTE
Caliente.COM

MEETING THE MEDIEVAL IN A DIGITAL WORLD

MEDIEVAL MEDIA CULTURES

Medieval Media Cultures offers analyses of how individuals interacted with written, visual, dramatic, and material media in medieval and early modern cultures, as well as how modern scholars interact with the remnants of medieval and early modern cultures via written, material, and now digital and electronic media.

This new series in media literacy welcomes proposals for monographs and essay collections in the fields of digital humanities, mapping, digital text analysis, games and gaming studies, literacy studies, and text production and interaction. We are especially interested in projects that demonstrate how digital methods and tools for research, preservation, and presentation influence the ways in which we interact with and understand these texts and media.

Series Editors

Toby Burrows, *University of Oxford*

Dorothy Kim, *Vassar College*

Richard Utz, *Georgia Institute of Technology*

Acquisitions Editor

Ilse Schweitzer VanDonkelaar, *East Lansing*

MEETING THE MEDIEVAL IN A DIGITAL WORLD

EDITED BY
MATTHEW EVAN DAVIS,
TAMSYN MAHONEY-STEEL, AND ECE TURNATOR

ARC HUMANITIES PRESS

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

© 2018, Arc Humanities Press, Leeds



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International Licence.

The authors assert their moral right to be identified as the authors of their part of this work.

Permission to use brief excerpts from this work in scholarly and educational works is hereby granted provided that the source is acknowledged. Any use of material in this work that is an exception or limitation covered by Article 5 of the European Union's Copyright Directive (2001/29/EC) or would be determined to be "fair use" under Section 107 of the U.S. Copyright Act September 2010 Page 2 or that satisfies the conditions specified in Section 108 of the U.S. Copyright Act (17 USC §108, as revised by P.L. 94-553) does not require the Publisher's permission.

ISBN: 9781641891929
e-ISBN: 9781641891936

<http://arc-humanities.org>

Printed and bound by CPI Group (UK) Ltd, Croydon, CR0 4YY

CONTENTS

List of Illustrations	vii
List of Contributors	xi
Introduction	1
MATTHEW EVAN DAVIS, TAMSYN MAHONEY-STEEL, AND ECE TURNATOR	
1 Statistical Analysis and the Boundaries of the Genre of Old English Prayer.	11
WILLIAM H. SMITH AND CHARLES L. BUTLER	
2 If (not “Quantize, Click, and Conclude”) {Digital Methods in Medieval Studies}	27
KATAYOUN TORABI	
3 Project Paradise: A Geo-Temporal Exhibit of the Hereford Map and <i>The Book of John Mandeville</i>	45
ALEXANDRA BOLINTINEANU	
4 Ghastly Vignettes: <i>Pierce the Ploughman’s Crede</i> , the Ghost of Shakespeare’s Blackfriars, and the Future of the Digital Past.	69
JIM KNOWLES	
5 Content is not Context: Radical Transparency and the Acknowledgement of Informational Palimpsests in Online Display	93
MATTHEW EVAN DAVIS	
6 Encoding and Decoding Machaut	123
TAMSYN MAHONEY-STEEL	
7 Of Dinosaurs and Dwarves: Moving on from <i>Mouvance</i> in Digital Editions	143
TIMOTHY L. STINSON	

8 Adam Scriveyn in Cyberspace: Loss, Labour, Ideology, and
Infrastructure in Interoperable Reuse of Digital Manuscript Metadata.157
BRIDGET WHEARTY

9 Digital Representations of the Provenance of Medieval Manuscripts.203
TOBY BURROWS

10 Bridging the Gap: Managing a Digital Medieval Initiative Across
Disciplines and Institutions223
JOSEPH KOIVISTO, LILLA KOPÁR, AND NANCY L. WICKER

Index241

LIST OF ILLUSTRATIONS

Figures

Figure 1.1. Dendrogram showing fingerprint word distribution in Old English poetry and prose “sampler” files.	19
Figure 1.2. Prayers and confessional materials using a combined 146-word fingerprint.....	21
Figure 1.3. Prayers and confessional materials using the thirty most frequent words in each set.....	22
Figure 1.4. Prayers, confessional materials, and the four “Confessional Prayers,” using a combined 146-word fingerprint.....	23
Figure 1.5. Prayers, confessional materials, and the four “Confessional Prayers,” using the thirty most frequent words.	24
Figure 1.6. Prayers, confessional materials, and the four “Confessional Prayers,” using the ten most frequent words.....	25
Figure 2.1. Combined <i>Beowulf</i> and <i>Blickling Homily XVII</i> 500-word chunks (Lexomics Tools).....	31
Figure 2.2. Dendrogram— <i>Canterbury Tales: General Prologue</i> portraits.....	37
Figure 2.3. Wife of Bath <i>General Prologue</i> —Word Cloud and Word Frequency List (generated using Voyant).....	38
Figure 2.4. Prioress <i>General Prologue</i> —Word Cloud and Word Frequency List (Voyant).....	38
Figure 2.5. Combined Wife of Bath and Prioress Portraits with function words present (Voyant Word Collocation tool).....	38
Figure 2.6. Combined Wife of Bath and Prioress Portraits excluding function words (Voyant Word Collocation tool).....	39
Figures 3.1 and 3.2. Hereford map, Earthly Paradise (covered by large circle) with rivers (indicated by small dots within the larger circle). ...	53
Figure 3.3. Hereford map, River from Paradise bears jewels into Prester John’s kingdom (triangle).....	55

Figure 3.4. Hereford map: Within the circle of Paradise, centre, flow its four rivers.	55
Figure 3.5. Hereford map: At the top, circled: Paradise; below, just above the lower border: the walled city of Jerusalem; above it, Christ's cross..	56
Figure 3.6. The Hereford map: Within the Earthly Paradise (encircled, above), the rivers stop before the wall. Outside the Earthly Paradise, the four rivers (from the top: Indus (Ganges), Tigris, Euphrates, and Nile) are not visually connected to Paradise.	58
Figure 3.7. The Hereford map: The triangle-studded line from the Cross (below) to Paradise and the angel (above) to the company of the blessed in the afterlife (top left).	61
Figure 4.1. Greyfriars second chapel, Phase I, 1244–48..	83
Figure 4.2. Greyfriars, Phase II, 1260–70.	84
Figure 4.3. Greyfriars, Phase III. Blackfriars is visible in the background.	84
Figure 4.4. Greyfriars, Phase IV, after 1270.	85
Figure 4.5. Greyfriars, Phase V, late fourteenth century..	86
Figure 4.6. Blackfriars second site, Phase I, 1236–45..	87
Figure 4.7. Blackfriars, Phase II (foreground). Greyfriars, Phase V (background)..	88
Figure 4.8. The underlying image is a digital copy of the Ralph Agas 1578 map of Oxford, a copy of which is in the Bodleian Museum.	88
Figure 5.1. The triadic relationship of Peirce's semiotic theory.	101
Figure 5.2. Infinite Semiosis—the Interpretant of the first sign becomes the Representamen of the next sign in the chain..	103
Figure 5.3. A constructed image intended to represent the opening showing fol. 300v–301r of the “Queen Mary Psalter.”	104
Figure 5.4. Fol. 300v and 301r as they appear on the British Library's Digitised Manuscripts site.	105
Figure 5.5. SQL Schema for the Minor Works of Lydgate website..	112

Figure 5.6. The <sourceDoc>-based encoding for the version of Lydgate's "Quis Dabit Meo Capiti Fontem Lacrimarum" in Jesus College Q.G.8, Jesus College, Cambridge, as rendered via the oXygen XML Editor.....	115
Figure 5.7. The author's master transcription file for the witness of the <i>Testament of John Lydgate</i> in Huntington HM140.	116
Figure 5.8. Model of the Clopton chantry chapel, Holy Trinity, Long Melford.	117
Figure 5.9. Force-directed model indicating the relationship between the three texts at Holy Trinity, Long Melford, in all witnesses.....	118
Figure 6.1. Screenshot of jechante.exeter.ac.uk showing the triplum text of the motet <i>Quant en moy / Amour et biauté / Amara valde</i> using the "Edition" tab and "Standard" view.....	126
Figure 6.2. TEI P5 conformant XML encoding of stanza one of "Helas! pour quoy virent onques mi oueil" (Lo53).....	126
Figure 6.3. Stanza one of "Helas! pour quoy virent onques mi oueil" (Lo53) as it appears in MS C.	127
Figure 6.4. Text and translation of stanza one "Helas! pour quoy virent onques mi oueil" (Lo53) without mark-up.	127
Figure 6.5. Example Exclusion File from the Roland Park Company Papers, Special Collections and Archives, Sheridans Libraries, Johns Hopkins University.....	130
Figure 6.6. Encoding with elements each having equal hierarchy.	131
Figure 6.7. Encoding with details nested under <person>.....	131
Figure 6.8. Rubric from the index of MS BnF fr 1584, f. Av.	134
Figure 7.1. Four views of the text available on the <i>Siege of Jerusalem</i> Electronic Archive.....	152
Figure 7.2. Apparatus of Bx, <i>Piers Plowman</i> Electronic Archive.	152
Figure 9.1. Example of multiple <provenance> elements in the TEI Guidelines.	206
Figure 9.2. Amphora of Tuthmosis III—an event-centric description (without aggregation and related digital representation entities).	214

Tables

Table 1.1.	Most frequent words in the <i>DOE</i> “Prayer” category (B12.4).....	15
Table 1.2.	Most frequent words in the <i>DOE</i> “Confessional Materials” category (B11).	15
Table 1.3.	Most frequent words in the <i>DOE</i> “Prayer” category with stop words removed.....	16
Table 1.4.	Most frequent words in the <i>DOE</i> “Confessional Materials” section with stop words removed.	16
Table 1.5.	Most frequent words in the four “Confessional Prayers” (B11.9.1, B11.9.3.1, B11.9.3.2, and B11.9.4).....	17
Table 8.1.	Comparison of XML-encoded metadata schemas employed by participating institutions.	177
Table 8.2.	Sample crosswalking of how the three example participating institutions mark up a work’s creator(s).	178
Table 8.3.	Sample crosswalk of Walters TEI to Stanford MODS.....	181
Table 8.4.	Descriptive metadata on the decorations and embellishments of Corpus Christi College Cambridge, Parker Library MS 304.....	189
Table 8.5.	Mock-up of what that macaronic description, drawing together English-language OAI:DC and French-language TEI, might look like....	193
Table 9.1.	The definition of the “Event” class in the Europeana Data Model.....	213
Table 9.2.	The components of a property graph for the purchase of Phillipps MS 16402.....	217
Table 9.3.	Sample nodes and relationships for Neo4j property graph.	217
Table 9.4.	Sub-objects of the “Manuscript” Object in the nodegoat framework....	217
Table 9.5.	A partial history of Phillipps MS 12264 modelled in nodegoat.....	218

LIST OF CONTRIBUTORS

ALEXANDRA BOLINTINEANU (alexandra.bolintineanu@utoronto.ca) is an Assistant Professor, Teaching Stream, in Medieval Digital Studies, at the University of Toronto. She is cross-appointed to the Centre for Medieval Studies and Woodsworth College. Her research interests include literary wonders in Old and Middle English; poetics of wonder in oral, literate, and digital environments; imaginary geographies; medieval and digital mapping; and digital humanities pedagogy.

TOBY BURROWS (toby.burrows@uwa.edu.au) is a Senior Researcher at the Oxford e-Research Centre in the University of Oxford, and a Senior Honorary Research Fellow in Humanities at the University of Western Australia. Current projects include “Mapping Manuscript Migrations” (funded by the Digging into Data Challenge) and “Collecting the West” (funded by the Australian Research Council). Between 2014 and 2016 he held a European Union Marie Curie International Incoming Fellowship in the Department of Digital Humanities at King’s College London. Publications include *Collecting the Past* (edited with Cynthia Johnston), forthcoming from Routledge.

CHARLES LEE BUTLER is currently the Director of Institutional Research and Assessment at Weatherford College. Trained in Statistics in the Sociology Department at Baylor University, his interest in the application of statistical methods has grown and diversified, ranging from text analysis in a variety of areas to Bayesian statistics in Educational Research. He continues to explore the power of R statistical programming language through research into variables associated with student success. Outside of his professional obligations, he explores applications of data analysis to the Digital Humanities.

MATTHEW EVAN DAVIS (matthew@matthewedavis.net) currently serves as a Postdoctoral Fellow with the Lewis and Ruth Sherman Centre for Digital Scholarship at McMaster University. Prior to this he was a Lindsey Young Visiting Faculty Fellow at the University of Tennessee’s Marco Institute and served as the Council for Library and Information Resources/Mellon Postdoctoral Fellow in Data Curation for Medieval Studies at North Carolina State University. He has published articles on medieval drama in *Theatre Notebook* and on the Lydgate verses at the Clopton chantry chapel in the *Journal of Medieval Religious Cultures*. He is the editor of the *Minor Works of John Lydgate* virtual archive (www.minorworksoflydgate.net) and is also working on a database charting the network of spiritual obligation captured in fifteenth-century wills and inventories.

JIM KNOWLES (jrknowle@ncsu.edu) is Teaching Associate Professor of English at North Carolina State University. He is Managing Editor of the *Piers Plowman Electronic Archive* and co-creator (with Michal Koszycki) of the *Oxford Friars Project*. He has

published articles on medieval literature and theology in the *Journal of Medieval and Early Modern Studies* and the *Yearbook of Langland Studies*. Together with Timothy Stinson, he is working on a hybrid print and electronic edition of Robert Crowley's three 1550 print versions of *Piers Plowman*.

LILLA KOPÁR (kopar@cua.edu) is Associate Professor of English and Director of the Center for Medieval and Byzantine Studies at The Catholic University of America in Washington, DC. She is an interdisciplinary scholar of the early Middle Ages with a special interest in the intersections of visual art and literature, myth and religion in the Viking Age, epigraphy and runology, and the art of commemoration. She has published on the iconography of Anglo-Saxon stone sculptures, on representations of Germanic myths and legends in early medieval northern Europe, and on runes and inscribed objects. Kopár is also co-director of the NEH-funded digital humanities initiative Project Andvari.

JOSEPH KOIVISTO (jkoivist@umd.edu) (MSLIS) is a systems librarian at the University of Maryland, serving the University System of Maryland and Affiliated Institutions (USMAI) consortium. Joseph earned his BA from the University of Scranton in 2009 and his MSLIS from the Catholic University of America in 2014. Additionally, he has worked for the District of Columbia Public Libraries and the Library of Congress. He has served as project assistant and project manager for Project Andvari, an NEH-funded research initiative, since 2014. Joseph's research interests include digital humanities work in libraries, critical perspectives on library technical infrastructure, crowdsourcing, and authority creation.

TAMSYN MAHONEY-STEEL (tamsyn@jhu.edu) is the Digital Scholarship Specialist for the Sheridan Libraries at Johns Hopkins University. She has a PhD in medieval French music and literature, and has published on medieval motets, linked open data, text mining the poetry of Guillaume de Machaut, library design, and collaborative scholarship. Her forthcoming monograph explores the scholarly benefits of using annotation tools on digitized medieval texts and manuscripts. Tamsyn teaches courses on digital humanities, using digital tools to explore medieval authors, and information literacy. At JHU she has instituted a series of seminars exploring the intersection of digital scholarship and issues of diversity and inclusion, and she has created workshops to introduce graduate students to text mining tools and methodologies.

WILLIAM SMITH (wsmith@wc.edu) is a Professor of English at Weatherford College, where he has taught classes in British Literature, Medieval World Literature, and Science Fiction Literature, among others. His research interests focus on Old English literature, especially the literature of learning and personal devotion, and on the application of statistical textual analysis methods to examinations of literary style.

TIMOTHY STINSON (tstinso@ncsu.edu) is associate professor of English at North Carolina State University. He has published articles on the Alliterative Revival, printing history, codicology, manuscript illumination, and the application of genetic analysis to the study of medieval parchment. He is editor of the *Siege of Jerusalem Electronic*

Archive, is co-founder and co-director of the Medieval Electronic Scholarship Alliance (MESA), associate director of the Advanced Research Consortium (ARC), co-director of the *Piers Plowman Electronic Archive*, and director of the Society for Early English and Norse Electronic Texts (SEENET). His research has received funding from the National Endowment for the Humanities, the Andrew W. Mellon Foundation, the Bibliographical Society of America, and the Council on Library and Information Resources.

KATAYOUN TORABI (torabik@tamu.edu) an instructional assistant professor in the Department of English at Texas A&M University. She is a specialist in Old and Middle English Literature, with strong research and teaching interests in the Digital Humanities. Her research, which incorporates both digital and traditional methodologies, investigates the connection between purgation and incarceration in pre- and post-mortem punitive spaces in Old and Middle English literature. Her work has been published in *Essays in Medieval Studies* and the *Association for Computing Machinery Press*.

ECE TURNATOR (turnator@mit.edu) received her PhD in medieval (Byzantine) history from Harvard University in 2013. Her dissertation is an interpretation of thirteenth-century Byzantine economy through an analysis of archaeological (coins and ceramics) and textual evidence. Between 2013 and 2016, she was a CLIR postdoctoral fellow at The University of Texas at Austin Libraries where she worked as coordinator for the Global Middle Ages Project, specialized in medieval data curation, studied and wrote about digital humanities and best practices for data curation and visualization, and taught in her areas of expertise. Currently, she is the humanities and digital scholarship librarian, and the liaison for history and linguistics departments, at the MIT Libraries.

BRIDGET WHEARTY (bwhearty@binghamton.edu) an Assistant Professor in English and Medieval Studies at Binghamton University (SUNY). She researches and publishes on the digitization of medieval manuscripts and on later medieval English literature and culture, especially the poetry of Geoffrey Chaucer, Thomas Hoccleve, and John Lydgate. From 2013–2015 she was a Council on Libraries and Information Resources (CLIR) Postdoctoral Fellow in Data Curation for Medieval Studies at Stanford University, where she served as data curator for a multi-institution digital manuscript index. She is co-editor of a special issue of *Archive Journal* dedicated to Digital Medieval Manuscript Cultures and is currently completing her first monograph—*The Digital Resurrection of Medieval Manuscripts*.

NANCY L. WICKER (nwicker@olemiss.edu) is Professor of Art History at The University of Mississippi. She has also been a Visiting Professor at Uppsala University and was elected to membership in the Royal Society of Humanities at Uppsala. Her research focuses on Scandinavian art during the Early Medieval Period, from the fifth century through the Viking Age, ca. 750–1100. She has had the support of fellowships from the National Humanities Center, the National Endowment for the Humanities, and the American Council of Learned Societies, as well as numerous grants. Wicker has served as Co-Director of Project Andvari, funded by NEH Digital Humanities Grants.

INTRODUCTION

MATTHEW EVAN DAVIS, TAMSYN MAHONEY-STEEL, AND ECE TURNATOR

SINCE THE ADVENT of the “New Philology,” medieval studies have been through an era of *renovatio*.¹ In the words of R. Howard Bloch, medievalists have confronted “tradition through renewal, rather than repeal.”² This time of renewal has become a key part of how we have dealt with the realm of the digital: its technologies, tools, and methodologies. In this book, we confront how those researching and teaching the Middle Ages traverse this realm, negotiating the “virtual divide” between the cultural artefacts that they study and the digital means by which they address those artefacts before presenting them to a variety of audiences. Part of this negotiation between digital and non-digital objects or subjects of study involves re-thinking and re-evaluating the ways that currently practicing medievalists develop their professional skills; this in turn will allow our community to reassess the ways in which the next generation of medievalists are educated. Although expressed through the perspective of the contributing medievalists, with their discipline-specific concerns, this book could thus also be perceived as a response to a rapidly evolving academic world. Rather than treating the digital humanities as its own discipline, the chapters in this volume examine it as a maturing methodology that increasingly intersects with field-specific curricula, boosted in turn by an upward surge in job searches looking for scholars who not only know digital tools, but who also, more importantly, understand technology as something other than an application.³

Thus, many of the articles in this book treat the “book form” as one of the successful technologies for delivering texts and images, along the lines argued by Geoffrey Rockwell and Stéfan Sinclair in *Hermeneutica*.⁴ Several directly tackle the issue of the importance of the medium of transmission, questioning whether we can ever separate text or image from the technology that presents or displays them. They characterize the inner workings of both the analogue and digital as transformative and interpretive methods of delivering a text or image each with their own unique affordances, and pose challenging

1 Nichols, “Introduction: Philology in a Manuscript Culture,” 1.

2 Bloch, “*New Philology Comes of Age*,” 1.

3 While we recognize that there is a rhetorical power in the declaration of field status, our belief is that the term “digital humanities” is too broad to be effective. Moreover, there are aspects of what is commonly referred to as “digital humanities”—GIS mapping, for example—that are not represented in this book. The capitalization used elsewhere also makes a claim regarding what digital humanities is and is not that we feel to be ultimately counterproductive. Individual authors within the volume, and the sources they cite, are making their own claims regarding the utility of the title and should be respected as such.

4 Rockwell and Sinclair, *Hermeneutica*, 161.

questions about the exegetical impact of any given medium. Just like the analogue, the digital is embedded and “materializes” in a context, is open to manipulation and exploration, and can fail in interesting ways.⁵ Each in its unique way, the essays in this book explore how digital contexts shape our interpretation of a text or image, taking into consideration both the benefits gained and the limitations thereof.⁶ In doing so, the authors are aware of the tension mentioned in *Hermeneutica* between the development process of a digital tool and writing as a research practice; they aim to avoid, as much as possible, the very real “temptation to interpret only through tools or only through discourse.”⁷ This book is thus a part of a trend towards increased formalization of methodologies in digital humanities and critical approaches toward the digital part of its name, which emphasizes an understanding of the work that goes into the development and design of digital tools and technologies in the first place.

The majority of the essays in this book underscore both the significance of having a technical understanding of the tools used to facilitate digital scholarship as well as the necessity of maintaining a critical distance regarding the ways that those tools intervene in our understanding of the text. Engaged with both digital praxis and critique of digital methodologies, they are influenced by debates about digital humanities that cast a critical look at how its typologies (humanities computing with its insistence on technical knowledge as “DH 1” versus “DH 2” as the typology that does not require technical knowledge—colloquially referred to as “hack” versus “yack”)⁸ developed, and the assumptions that may lurk under such a divide. Adeline Koh’s call for an “expanded genealogy” for the digital humanities that rests on an inclusive and critical formulation of its history, brings into focus the issues of inequality, race, gender, class, ableism, and insists on an opening toward a non-western focused, truly global studies, is a case in point.⁹ Likewise, although none of the essays in this book directly undertake socio-economic criticism of globally networked computational economies we recognize the invaluable, healthy discussions that take place in the areas of critical code studies as well as criticisms of socio-economic networks that rest on the digital’s feigned objective rationality.¹⁰ David Berry emphasizes the “need to explore the historical, philosophical,

⁵ Leonardi, “Digital Materiality? How Artifacts Without Matter, Matter,” argues that it is the uses of the digital that make it material. Even though it may not have physical matter, the digital “materializes” in the ways it is used and applied.

⁶ *Hermeneutica*, 166.

⁷ *Hermeneutica*, 166.

⁸ Ramsay’s article, although no longer available (itself an example of the type of problem many of the articles referenced herein hope to avoid), is an essential source for this distinction. See also Nowviskie, “On the Origin of ‘Hack’ and ‘Yack,’” in the 2016 edition of *Debates in the Digital Humanities* (which is taken from a blog post of the same title), and Warwick, “Building Theories or Theories of Building?”

⁹ Koh, “Niceness, Building, and Opening the Genealogy of the Digital Humanities.”

¹⁰ Zehle and Rossiter, “Meditations of Labor”; Rossiter, “Materialities of Software: Logistics, Labour, Infrastructure [Extended Version]”; Berry and Fagerjord, *Digital Humanities. Knowledge and Critique in a Digital Age*.

theoretical and critical context for particular kinds of the various forms of digital *praxis*,” and it is in this spirit that we hope to provide not just concrete examples of ongoing projects for the medievalist interested in digital scholarship, but a primer in ways to explore the intersection of the digital on topics discussed in these pages.¹¹ We are confident that both seasoned digital medievalists and all interested in digital humanities will find the articles in this book approachable and deem them as valuable contributions to the field of digital medieval studies.

As such, this book signifies not just a collection of tools and methods, but also a move toward values expressed in both the theory and practice of digital humanities, a move that takes its cues from a pedagogy-focused article by Sean McCarthy and Andrew Witmer.¹² In many chapters the underlying emphasis is on the importance of learning holistically: the value of teaching students to understand and use critically the technology that lies “under the hood” not only as a tool or a means to an end, but as a fundamental part of a scholar’s critical apparatus; in short, to move beyond sheer quantification of data and clicking on buttons.¹³ It is also worth emphasizing that even though students may not want to use the technologies or apply the methodologies mentioned in this volume in their own work, others in their field may. Therefore, they should be prepared to assess and respond intelligently to their colleagues’ research, and as digital tools and methodologies become more integrated into the traditional research practices it is becoming increasingly necessary to teach students these skills as part of their training—even if they may in fact end up not using them. It is worth underscoring in this context that several of the essays in the book speak to the importance of combining the use of digital tools and methodologies with traditional close reading techniques.¹⁴ Others explore the meaning of the unique physicalities of the medieval manuscript and how they both are reflected in and differ from its digital analogue.¹⁵ Further, within the framework of digital humanities, the book covers a host of significant issues that the academy and GLAM (galleries, libraries, archives, and museums) institutions face together such as differences in models of information organization, metadata standards,

11 Berry, *Critical Theory and the Digital*, 209.

12 McCarthy and Witmer, “Notes Towards a Values-Driven Framework,” 2016.

13 On the importance of making the assumptions and editorial decisions explicit in the user interface, see, Ruecker, “Interface as Mediating Actor,” 397–407. On programming as a way of thinking, see Montfort, *Exploratory Programming*, 8. In this context, it is equally worth mentioning Berry’s approach to programming beyond “a *skill* for a new economy,” as a means to understand from the inside out the socio-economic hierarchies created by code: Berry, *Critical Theory*, 209.

14 Many of the articles, especially Van Zundert’s conclusion in *Analysis of Ancient and Medieval Texts and Manuscripts*, edited by Andrews and Macé, constitute important precedents. “The Digital Middle Ages: An Introduction,” edited by Birnbaum, Bonde, and Kestemont, was published shortly after the articles for this book were submitted for publication. It also has relevant, valuable contributions.

15 We would be remiss at this point if we did not mention McGann and Shep’s important contributions to these discussions in *A New Companion to Digital Humanities*.

and the reduction in quality, or lossiness, of the connections, commonly referred to as “crosswalks,” built between those standards.

That medievalists are at the forefront of this trend should not be surprising, considering that they have pioneered much of the methodological application of technology in their work, especially in the fields of digitization and textual analysis.¹⁶ As such, this book looks at the intersection between digital humanities, in its many forms, and medieval studies in its equally myriad affordances from the point of view of medievalists who have created digital resources or applied digital tools and methodologies in their scholarship. Text encoding and analysis, data modelling and provenance, and 3D design are all discussed as they apply to western European medieval literature, history, art history, and architecture. Practically speaking, it does so through an examination of four broad topics: stylometric analysis, in particular the version of stylometry championed by Michael D. C. Drout under the term “lexomics,” the intersections of place and space in the analysis of texts, the presentation and display of digital or virtual facsimiles of medieval manuscript texts, and questions of the infrastructure development and project management that underlie any digital project. The chapters themselves, while falling into one of these four categories, often take cues from or speak to elements of the others, so that the book as a whole provides an example of the state of much of medieval studies as practiced digitally today.

William Smith and Charles Butler’s chapter begins the volume by applying computational analysis not in an examination of word usage, but instead to attempt to determine whether a text is poetry or prose. By running lexomic analyses on texts classified as either “Confessional Materials” or “Prayer” in Neil Ker’s *Catalogue of Manuscripts Containing Anglo-Saxon*, the authors hope to identify whether such computationally assisted methods can help to identify genre in a specific genre of prose or poetry. They do so by creating dendrograms, using the statistical software package R, to show the distribution of various previously identified texts. Smith and Butler observe that, using the cluster analysis algorithm, they could detect whether a text was poetry or prose after an initial identification of representative words—“fingerprint words”—one set for the poetry dataset and one for the prose dataset.

Katayoun Torabi, as the title of her chapter suggests, notes that the use of digital tools and methodologies cannot replace but instead must stem from traditional humanities scholarship. By pairing traditional humanistic analysis with computer-assisted textual analysis using the Lexomics and Voyant Tools software suites in the examination of Old and Middle English texts such as *Beowulf*, the *Blickling Homilies*, and the *Canterbury Tales*, Torabi shows that the digital method can refine, affirm, or deny hypotheses, assist in the interpretation of texts, and aid the researcher in discovering patterns of word usage that otherwise remain undetectable. Through her examination of this work, she notes that researchers cannot simply expect the tool to produce results, but must understand computational tools critically, which requires structured training within the field.

¹⁶ See, for example, the introduction by Porter, “Medievalists and the Digital Scholarly Edition.”

In her chapter, Alexandra Bolintineanu discusses the development of a digital comparison tool, entitled Project Paradise, which maps and annotates two late-medieval representations of the Earthly Paradise: the thirteenth-century Hereford mappa mundi (world map) and the fourteenth-century travel memoir *The Book of John Mandeville*. The article focuses specifically on the version of *The Book of John Mandeville* that appears in a British Library manuscript (Royal 17 C. xxxvii). This so-called Defective Version was the most widespread in England. Project Paradise's digital comparison enables Bolintineanu to project textual evidence from the Defective Version against the Hereford map, creating a geo-temporal exhibit that places the mental geography of Mandeville's world on a digital representation of a near-contemporary map of that selfsame geography. This juxtaposition of the two material objects creates a space for Bolintineanu to examine Mandeville's "spatial poetics against close-grained textual and visual evidence." Each Mandeville reference that Bolintineanu notes tracks three strands: first, the Middle English text from British Library manuscript; second, the Old French text from the *Livre de merveilles*, in Christiane Deluz's critical edition, and third, a translation of the Middle English into modern English. Organizing the digital tool in this way allows readers to focus their attention on the British Library manuscript, but at the same time to compare its paradise references to the continental tradition as represented by the *Livre*. The process of developing Project Paradise, argues Bolintineanu, granted insights regarding the creation process of maps such as the Hereford mappa mundi as a series of negotiations and confluences between text-based and spatial narratives.

Jim Knowles's chapter utilizes a combination of traditional scholarship and new digital methodologies to upend the received wisdom that *Pierce the Plowman's Crede* is situated in the urban streetscape of fourteenth-century London. Based first on an analysis of the original evidence linking the poem to London, Knowles notes that the analysis that supported this conclusion is the result of early twentieth-century architectural history intended not to find the location of the fourteenth-century poem, but Shakespeare's sixteenth-century indoor theatre at Blackfriars. What was meant to be a description of the friary is instead taken to be *the* identification of London with the poem by subsequent American Shakespeareans. Having established the dubious provenance of the scholarly argument for a London location, Knowles then connects this scholarship to the three-dimensional models of lost Franciscan and Dominican mendicant houses he created alongside Michal Koszycki in order to underscore what he sees as a "tension ... between the seductive power of technology-driven reconstruction of the past in all its detail ... and a deep discipline-driven scepticism and resistance toward the spectre of empiricism that such projects conjure up in the minds of (some) literary critics." Instead of treating his models as a finished product, he addresses the idea of building digital projects as a thought experiment, which allows for insights that might not have been available otherwise. Furthermore, the models allow students to seek insights, based on a series of guided questions, rather than simply existing as a "reified visible model"—a virtual facsimile treated as though it existed in the physical world.

Matthew Evan Davis's chapter introduces a group of interrelated pieces examining the connections between digital, web-based archives of manuscripts and their physical originals. Davis argues for the importance of the interaction between form and

content: the ways in which content is presented can fundamentally alter a reader's understanding and perception of it. Therefore, rather than simply being a means to an end, without theoretical influence, the platform by which content is delivered is significant. Davis examines this through an analysis of the evolution of the file/folder metaphor in computing, its uptake and transition into the digital world, and the impact of that uptake on both popular and scholarly thinking. As a scholar of medieval literature familiar with digital tools and platforms as well as their histories, Davis moves in his chapter from the general examination of the file/folder metaphor into a description of his work in creating the Minor Works of John Lydgate virtual archive. While other authors—most notably Koivisto, Kopár, and Wicker in the final chapter—emphasize the vast gap between technologists and medievalists on a project level, Davis's chapter underscores the significance of the shortfall in training, and further underscores the need to realize two things: first, that both technologists and scholars make assumptions about the others' areas of expertise that can prove erroneous, and second, that the act of creating digital tools is itself an act of theory-making that influences the direction of the overall project. Scholars, researchers, and instructors should be aware of the implications of both sets of assumptions, and begin to master a level of awareness that will allow them both to work more easily with technologists in articulating their shared vision for a project and to better appreciate the impacts of technology on our interpretation of material artefacts as cultural heritage items.

In her "Encoding and Decoding Machaut," Tamsyn Mahoney-Steel draws attention to the unique existence of the digital edition and highlights its capacity to free the editorial process from the restraints of traditional codex-based editions. Based on her experience working on the TEI-based XML encoding for *Je Chante Ung Chant*, a web-based archive of largely fourteenth-century songs and motets, Mahoney-Steel relinquishes the purported ideal of the digital's capacity for capturing the physical as well as the ideal of creating an edition exactly like the physical in verisimilitude. Instead, she argues that the online version of a text has a life and form of its own that is both unlike and somewhat independent of the material object it claims to "replicate." Not unlike Davis's chapter in the book, Mahoney-Steel writes that a digital edition can, with all the available transcribed variants and their associated manuscript images, appear more holistic than a print edition. However, it cannot claim to capture its sources fully even in an expertly encoded version—the act of moving the data to a new realm changes the context and creates a unique, digital adaptation of the material text. This digital adaptation has a life of its own, which makes the process of encoding art, not craft.

In his chapter, Timothy Stinson considers the concepts of *variance* and *mouvance*, as interpreted by book historians such as Bernard Cerquiglini, Paul Zumthor, and Stephen Nichols in their work editing medieval French texts. Similarly to Mahoney-Steel, Stinson argues that the digital captures the fluidity (*mouvance*) of the text better than the print version, based in part on his work with the *Piers Plowman* Electronic Archive (PPEA) and the *Siege of Jerusalem* Electronic Archive (SJEA). While most, if not all, medieval texts pose manifold complexities to their editors, Stinson argues that "they are in no way as complex as the surviving corpus of *Piers* texts and the editorial cruces that they famously pose," as there are a large number of surviving *Piers* manuscripts with

significant differences in readings as well as various other challenges and eccentricities. He further posits that the collaboratively composed, image-rich, and often anonymous hypertext pages that make up the web have qualities similar to densely glossed and illuminated folios of medieval manuscript culture. Similar to medieval manuscripts, digital editions of texts are therefore inherently fluid in communicating various editorial approaches and in their treatment of variants, or *variance*.

Bridget Whearty's chapter exposes the diligent work necessary to create the descriptive information, or metadata, for medieval manuscripts and other cultural heritage items held in various collections—each with its own preferred nomenclature. The first part of the article focuses on her work on the Digital Manuscript Index, “a multi-institution experiment working toward that dream of a single search portal,” describing how copying and transformation inevitably creates change and sometimes loss in making analogue medieval manuscripts understandable by a computer as well as a researcher. The second section of the article is a detailed analysis of the Corpus Christi College and Stanford University Library's Parker Library on the Web project, honing in on the challenges of reusing data and associated metadata from older projects often made according to former “best practices” that might no longer apply. The final section of the article focuses on the shortcomings of multilingualism in digital humanities when it comes to describing terminology that must be precisely explicated to be understood across various computational platforms and in different institutional environments. Through an examination of the issues involving transforming the Walters Art Museum's TEI-encoded XML into Stanford's local implementation of Metadata Object Description Schema, or MODS, encoded XML, Whearty shows the complications that arise from the encoding of different analogue standards in constituting new digital platforms. The lessons provided by this work show how the needs of institutional partners are foregrounded in the creation of crosswalks, and how in doing so important information can be lost.

In his chapter, Toby Burrows also examines issues regarding metadata with a specific emphasis on representations of provenance in the collection of Sir Thomas Phillips, a now-dispersed private collection of over 40,000 items. He notes that even though manuscript provenance is often included in the manuscript extensions of Machine-Readable Cataloging, or MARC, record metadata for the manuscript, this information is not represented in standard MARC-based online displays. Some online catalogues, such as Harvard University's HOLLIS catalogue, and major research libraries have devised ways to make the provenance information accessible on these standard displays via workarounds such as the inclusion of the term “former owner” separated from the owner's name via a comma—practices that vary from institution to institution. MARC's (possible) successor, the International Federation of Library Associations and Institutions-developed FRBR (Functional Requirements for Bibliographic Records) also fails to provide specific modelling for provenance data. After examining the practices of a number of institutions with large manuscript holdings, Burrows turns his attention to the cultural heritage institutions that contribute to CIDOC (Comité International pour la Documentation Conseil International des Musées) and its data model, the Conceptual Reference Model (CRM), which is used by Europeana. After noting the gaps in the

FRBR/CRM data model, Burrows suggests the use of additional tools, such as Neo4j and nodegoat, to overcome some of the issues in encoding the history of provenance of manuscripts.

As Burrows and Whearty have shown, “translation” between different cataloguing and encoding practices can be a rather complex and at times defective matter. Joseph Koivisto, Lilla Kopár, and Nancy Wicker focus in the final chapter on the complications of “translating” field-specific expertise in project-based collaborations among medievalists, information scientists, librarians, and technologists.¹⁷ The goal of the NEH-funded Project Andvari is to provide an “integrated access to collections of northern European art and artefacts of the early medieval period (fourth to twelfth centuries),”¹⁸ modelled on linked data (LD) architecture. When they first began collaborating, Koivisto was an information school student (currently a systems librarian); Kopár was and remains a professor of medieval literature and historical linguistics, and Wicker was and remains a professor of medieval art history. Other collaborators include the Institute for Advanced Technologies in the Humanities (IATH) at the University of Virginia, as well as individuals in partnering academic and cultural heritage institutions—the so-called GLAM institutions. As Koivisto, Kopár, and Wicker attest in their article, even though rewarding for all involved areas of expertise, the lack of formal training in project development and management skills has the potential to negatively implement the proliferation of valuable digital medieval projects and initiatives in the future.

The second and current stage of Project Andvari aims to produce a functioning platform using data from the British Museum, the Swedish National Heritage Board’s Kringla database, and the Norwich Castle Museum. The paper underscores the need to develop new models of skills and values training for field experts and practitioners in order to advance scholarship, and suggests that Project Andvari can serve as a model for such work. Project Andvari’s collaboration across multiple institutions by individuals with wide and varied sets of expertise poses welcome challenges, requiring the design of physical spaces for these individuals to meet, new methods for training students and faculty, and mechanisms of recognition for scholars who engage with the tools provided. These models, and the ideas incorporated within, have applications well beyond medieval studies.

As evidenced by these chapters, *Meeting the Medieval in a Digital World* reinforces the fact that medievalists have traditionally been at the forefront of digital scholarship, particularly when it comes to the digitization, transcription, and dissemination of virtual facsimiles and cultural heritage items. More importantly, this book shows that the perceived affiliation—where medieval studies is viewed as a source of raw material or an area of application for the field of digital scholarship—operates in the opposite direction. Digital scholarship is the handmaiden of the subject it serves; it is a set of tools and methodologies and a means of interaction that allows scholars to approach their

¹⁷ For a recently published overview of project management among digital humanities practitioners see Siemens, “Project Management and the Digital Humanist,” 343–57.

¹⁸ “About Project Andvari,” www.andvari.org/index.php.

material in innovative ways. General books on digital humanities tend to assume a set of circumstances and relationships between researcher, textual corpora, and physical objects that is not universally the case. Medievalists frequently do not have a representative set of data to mine, as might be the case for scholars of the post-print era, with access to standardized and reproduced texts. Instead, the discussion in this volume is informed by a sensitivity to the evidence of objects, their contexts, and the relationships between them. We hope that this book will broaden the horizons of medievalists considering entering or already happily ensconced in the digital world, while asking digital humanists to challenge their assumptions about the contexts of our objects of study and the tools and methodologies used to approach them.

Bibliography

- Andrews, Tara, and Caroline Macé, eds. *Analysis of Ancient and Medieval Texts and Manuscripts: Digital Approaches*. Turnhout: Brepols, 2014.
- Berry, M. David. *Critical Theory and the Digital*. New York: Bloomsbury, 2014.
- Berry, M. David, and Anders Fagerjord. *Digital Humanities. Knowledge and Critique in a Digital Age*. Cambridge: Polity Press, 2017.
- Bloch, R. Howard. "The New Philology Comes of Age." In *Rethinking the New Medievalism*, edited by R. Howard Bloch, Alison Calhoun, Jacqueline Cerquiglini-Toulet, Joachim Küpper, and Jeanette Patterson, 1–11. Baltimore: Johns Hopkins University Press, 2014.
- Birnbaum, David J., Sheila Bonde, and Mike Kestemont. "The Digital Middle Ages: An Introduction." Special issue, *Speculum* 92 (2017). www.journals.uchicago.edu/doi/abs/10.1086/694236
- Koh, Adeline. "Niceness, Building, and Opening the Genealogy of the Digital Humanities: Beyond the Social Contract of Humanities Computing." *Differences. A Journal of Feminist Cultural Studies* (2014). <http://differences.dukejournals.org/content/25/1/93.full.pdf+html>.
- Leonardi, Paul. "Digital Materiality? How Artifacts Without Matter, Matter." *First Monday* 15 (2010). <http://dx.doi.org/10.5210/fm.v15i6.3036>.
- McCarthy, Sean, and Witmer, Andrew. "Notes towards a Values-Driven Framework for Digital Humanities Pedagogy." *Digital Pedagogy Lab* (March 29, 2016). www.digitalpedagogylab.com/hybridped/values-driven-framework-digital-humanities-pedagogy/.
- McGann, Jerome. "Marking Texts of Many Dimensions." In *A New Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens, John Unsworth, 358–76. Chichester: Wiley Blackwell, 2016.
- Montfort, Nick. *Exploratory Programming for the Arts and Humanities*. Cambridge, MA: MIT Press, 2016.
- Nichols, Stephen G. "Introduction: Philology in a Manuscript Culture," in "The New Philology." Special issue, *Speculum* 65 (1990): 1–10.
- Nowvskie, Bethany. "On the Origin of 'Hack' and 'Yack'." In *Debates in the Digital Humanities*, edited by Matthew Gold. New York: CUNY Press. <http://dhdebates.gc.cuny.edu/debates/text/58>

- . “On the Origin of ‘Hack’ and ‘Yack’” (blog), January 8, 2014. <http://nowviskie.org/2014/on-the-origin-of-hack-and-yack/>
- Porter, Dot. “Medievalists and the Digital Scholarly Edition.” *Annual of the Documentation for Scholarly Editing* 34 (2013). www.scholarlyediting.org/2013/essays/essay-porter.html. See also Project Andvari. www.andvari.org/index.php
- Rockwell, Geoffrey, and Stéfan Sinclair. *Hermeneutica. Computer-Assisted Interpretation in the Humanities*. Cambridge, MA: MIT Press, 2016.
- Rossiter, Ned. “Materialities of Software: Logistics, Labour, Infrastructure [Extended Version].” *Organized Networks*, February 13, 2014. <http://nedrossiter.org/?p=387>.
- Ruecker, Stan. “Interface as Mediating Actor for Collection Access, Text Analysis, and Experimentation.” In *A New Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens, and John Unsworth, 397–407. Chichester: Wiley Blackwell, 2016.
- Shep, Sydney J. “Digital Materiality.” In *A New Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens, and John Unsworth, 322–30. Chichester: Wiley Blackwell, 2016.
- Siemens, Lynne. “Project Management and the Digital Humanist.” In *Doing Digital Humanities. Practice, Training, Research*, edited by Constance Crompton, Richard J. Lane, and Ray Siemens, 343–57. New York: Routledge, 2016.
- Warwick, Claire. “Building Theories or Theories of Building? A Tension at the Heart of Digital Humanities.” In *A New Companion to Digital Humanities*, edited by Susan Schreibman, Raymond Siemens, and John Unsworth, 538–52. Chichester: Wiley Blackwell, 2016.
- Zehle, Soenke, and Ned Rossiter. “Meditations of Labor: Algorithmic Architectures, Logistical Media, and the Rise of Black Box Politics.” *Organized Networks*, January 5, 2015. <http://nedrossiter.org/?p=453>.

Chapter I

STATISTICAL ANALYSIS AND THE BOUNDARIES OF THE GENRE OF OLD ENGLISH PRAYER

WILLIAM H. SMITH AND CHARLES L. BUTLER

ONE OF THE most basic questions in literary analysis, the question of the genre of the text being examined, when applied to the study of Old English literature is also one of the most fraught. Though obviously many texts fall clearly into recognized categories, the corpus of Old English literature also contains many examples of texts that blur the lines between genres. The most obvious case is the line between poetry and prose in Old English, which remains the matter of much scholarly debate.¹ Even within those larger genres, however, the boundaries of more specific textual categories can often be difficult to determine,² raising hard questions about the way both individual texts and groups of texts may have been understood by Anglo-Saxon authors and readers.

Traditional methods of scholarship can offer some answers to those questions, of course. Textual details—the presence of poetic vocabulary and meter in an Old English poem, the use of traditional sermon structure and direct address to an audience in an Old English homily, for example—usually provide some suggestion as to the intended genre of Old English texts. Manuscript context often offers us further clues to the way an individual text may have been perceived. But especially in cases where such details are not clear, it may be useful to employ digital tools that can examine large groups of texts quickly and perhaps discern patterns and differences that would go unnoticed in a more traditional approach.

Michael Drout's work in statistical analysis of Old English poems, an approach that he has termed "lexomics," has demonstrated the effectiveness of digital tools in detecting subtle divisions within an individual text as well as lines of influence between one text and another.³ By examining relative word frequency, Drout's algorithm has successfully recognized the internal division between *Genesis A* and *Genesis B* and identified the portion of the Old English poem *Daniel* that is related to the poem *Azarias*.⁴ Although lexomic methods have been used primarily to identify relationships of sections inside single texts, similar analysis of word frequency has long been used as well to help determine authorship of contested or anonymous works.⁵ This approach has not so far been

¹ For only one notable and recent entry into this long-standing debate, see Bredehoft, *Authors, Audiences, and Old English Verse*.

² See, for example, Battles, "Toward a Theory of Old English Poetic Genres."

³ Drout et al., "Of Dendrogrammatology." The fullest treatment of the lexomic methodology is found in Drout, *Tradition and Influence in Anglo-Saxon Literature*.

⁴ Drout et al., "Of Dendrogrammatology," 307–19.

⁵ See, for example, Burrows, "Questions of Authorship: Attribution and Beyond," and Hoover, "Testing Burrows's Delta." The most significant application of this approach to Old English literature is Gill, Swartz, and Treschow, "A Stylometric Analysis of King Alfred's Literary Works."

applied to questions of genre differences in Old English literature, but it is not unreasonable to suggest that the same kinds of word use patterns that relate to source influence or authorial style might also be discernible in specific genres.⁶

The data at the heart of approaches such as lexicomics is word frequency. While genre definitions certainly cannot be reduced to simple differences in vocabulary, there's little question that the use of specific groups of words and, more importantly, patterns of word use are often common among texts within a specific genre. In Old English, for example, we know that poetic texts often contain markedly different vocabulary when compared to prose texts. Scholars familiar with those differences in vocabulary would likely be able to identify a previously unknown Old English text as an example of either prose or poetry based at least in part on the appearance or non-appearance of individual words. If provided with a list of typically poetic or prosaic words, a computer could do the same, of course. But can a computer recognize the differences between Old English prose and Old English poetry without a predefined lexicon? In other words, are the differences in word use patterns between poetry and prose distinct enough for a computer to recognize those differences and establish reliable boundaries between prose texts and poetic texts? And what happens when we ask that same computer to examine the boundaries around a more specific textual genre, such as Old English prose prayer?

The only systematic attempt to define a corpus of Old English prayers is that represented by the list of texts in the *Dictionary of Old English (DOE)* corpus of Old English.⁷ Perhaps systematic is the wrong word to use here, however, since the *DOE* offers no rationale for including or excluding an individual text from the category of prayer. That category, represented by Cameron number B12.4, includes only twenty-four individual prayers (technically twenty-five, since two alternate versions of the Pater Noster are included).⁸ Some texts that would appear, based on their textual forms and manuscript context, to be prayers are not included in the *DOE* list. Most intriguing among these omissions are four texts included instead in the category of Confessional and Penitential Texts, Cameron number B11.⁹ The omission of these texts from the category of prayers is especially interesting because all four are generally treated as prayers in the editions used by the *DOE* in compiling its Old English corpus.¹⁰ In classifying

⁶ The application of statistical analysis tools to the question of genre boundaries is not unknown. See, for example, Craig, "Stylistic Analysis and Authorship Studies." For a somewhat similar experiment to our own, using nineteenth-century British novels as a dataset, see Allison, "Quantitative Formalism: An Experiment."

⁷ Frank and Cameron, *A Plan for the Dictionary of Old English*.

⁸ Since the original Cameron *Plan* was published in 1973, one additional Old English prayer has been found and added to that number (as Cameron Text B12.4.12). See Rushforth, "The Barrow Knight."

⁹ The four texts in question are described as "Forms of Confession and Absolution" and are represented by Cameron numbers B11.9.1, B11.9.3.1, B11.9.3.2, and B11.9.4.

¹⁰ The texts known as B11.9.1 and B11.9.4 are edited by Logeman, "Anglo-Saxonica Minora." The texts numbered as B11.9.3.1 and B11.9.3.2 have been edited most recently by Pulsanio and McGowan, "Four Unedited Prayers."

the texts in this way, the *DOE* is clearly following Neil Ker's *Catalogue of Manuscripts Containing Anglo-Saxon*, which indexes all four under the heading "Forms of Confession," rather than under "Prayer."¹¹ Certainly all four texts contain confessional elements.¹² The question remains, however, whether those elements are pervasive or significant enough to justify classifying these texts as confessional materials rather than prayers. After all, many Old English prayers are confessional in nature. Within its own category of prayer, in fact, the *DOE* describes five texts specifically as "confessional prayers,"¹³ begging the question of what separates a "confessional prayer" from a "form of confession," if the latter is clearly presented and structured as a prayer. More importantly, this division raises the question of where the boundary between prayers and confessional materials should be drawn.

Statistical analysis techniques similar to Drout's lexomics approach, which make use of word frequency distribution patterns, can help draw these boundaries, although the success of the model depends largely on the specific parameters of the methodology. The number and type of words counted, for example, is a significant determinant of the ability of the approach to differentiate groups of texts. Most statistical analyses of the style or authorship of literary texts—often referred to as computational stylistics or stylometrics—restrict their examinations to a small group of the highest-frequency words, usually function words such as articles, conjunctions, and prepositions. The rationale for doing so is that markers of authorial style should be independent of the content of any individual text, so content words that might be influenced by the nature of the specific work should not be included in the analysis. In a 2007 stylometric study of the authorship of the Old English Prose Psalter, for example, statistical determination was made on the basis of only seventeen commonly occurring words, almost all of them prepositions, such as *mid*, *of*, and *to*, or simple verbs like *is* and *wæs*.¹⁴ On the other hand, Michael Drout's lexomic approach is based on an analysis of the frequency of all words appearing in a text. For the purposes of finding subtle divisions inside a long text, this model works well, but it raises problems for genre identification. Using again the example of Old English poetry and prose, although poetic words may well serve as a visible marker of the genre of poetry, those words usually make up only a small percentage of the total words in any poem.¹⁵ Thus, comparing prose and poetry simply on the basis

11 Ker, *Catalogue of Manuscripts Containing Anglo-Saxon*.

12 All four texts contain explicit references to confession using the verb *andettan*, for example. But that verb appears in a majority of the texts identified as "prayers" as well.

13 Specifically, these texts are Cameron numbers B12.4.3.1–B12.4.3.5.

14 Gill, Swartz, and Treschow, "A Stylometric Analysis of King Alfred's Literary Works." The authors of this study argue that their findings demonstrate conclusively that King Alfred is not the author of the Prose Psalms. Janet Bately has raised several objections to their approach, however, among which is the notion that even function word usage may be affected by the specific content of a text, especially in the case of a translation. See Bately, "Did King Alfred Actually Translate Anything?"

15 In the sample datasets of Old English poetry and prose described later in the essay, for example, approximately 20 per cent of the words appearing in the poetry set were exclusive to that set. Obviously, some portion of that variation can be written off as incidental to the specific texts chosen,

of overall word counts may not result in a very significant separation between the two groups. Similarly, counting only the highest-frequency function words may make sense when trying to identify authorial style independent of textual content, but the frequency of these kinds of words may not vary much from one genre to the next.

Because we are interested specifically in the dispersion of words due to genre differences, it makes the most sense to focus on those words that are most affected by the genre of the text. But which words are those? In the case of Old English poetry, we could restrict the count to only those words that have been identified as exclusively poetic, but doing so would make any comparison to non-poetic texts meaningless. Furthermore, it stands to reason that the existence of poetic vocabulary is not the only lexical difference between poetry and prose.

When we turn our examination to a more specific genre, such as prayer, it becomes quite a bit more difficult to identify any static group of genre-specific words. Verbs like *biddan* (to pray) and *forgifan* (to forgive) certainly show up frequently in devotional texts, but they're hardly exclusive to those texts, nor are they common enough in Old English prayers to be characterized as necessary elements. The fact that no identifiable group of "prayer words" exists does not mean, however, that there is no such thing as "prayer language." Old English prayers are, for instance, relatively formulaic in style, using common phrasing often taken from Latin liturgical models.¹⁶ They also share content that is likely to be reflected in the distribution of common words, though perhaps not in only the very most common function words.

In order to identify the most frequent words in Old English prayers, we created a dataset that consisted of all of the texts categorized as "Prayers" by the *DOE*. Using the machine-readable corpus of Old English created by the *DOE*, we compiled all texts in category B12.4 into a single file and then scrubbed them of all formatting and special characters, using a simple Python script. For the purposes of clarity, all appearances of *eth* (ð) and *thorn* (þ) were folded together. Likewise, *and*, *ond*, and the Tironian note (ꝛ) were folded into a single morpheme (*and*).¹⁷ We then followed the same steps to create a dataset of the texts categorized by the *DOE* as "Confessional Materials" (B.11). Because of the contested nature of the four "confessional prayers" (B11.9.1, B11.9.3.1, B11.9.3.2, and B11.9.4), they were not included in this file but were placed in their own dataset, with the same procedures applied.¹⁸

Table 1.1 lists the ten most common words in all of the texts included in the *DOE* category of "Prayer." A glance at this list is enough to see that only a few of the most common words in these texts are the product of genre identity. The prevalence of first-person pronouns (*ic*, *me*, *min*) and references to God (*drihten* and, probably, *þu*) are common

so it seems safe to assume that fewer than 20 per cent of the words appearing in Old English poetry are "poetic" words.

16 See Smith, "The Tradition of Vernacular Prayer in Anglo-Saxon England."

17 For an alternate take on these decisions, see Drout et al., "Of Dendrogrammatology," 304.

18 It should be noted, however, that those four texts amount to just over 2,000 words and would likely have very little effect on the word dispersion in a category comprising almost 25,000 words total.

Table 1.1. Most frequent words in the *DOE* “Prayer” category (B12.4).

Rank	Word	Relative Frequency
1	And	.075
2	þe	.038
3	On	.030
4	Ic	.027
5	Me	.024
6	þæt	.022
7	þu	.021
8	Drihten	.015
9	To	.014
10	Min	.014

Table 1.2. Most frequent words in the *DOE* “Confessional Materials” category (B11).

Rank	Word	Relative Frequency
1	And	.069
2	On	.027
3	þe	.023
4	He	.021
5	þæt	.020
6	His	.018
7	Man	.015
8	To	.014
9	Gif	.014
10	Ne	.014

elements in devotional texts, but the remaining words on the list are purely functional in nature and likely have little connection to the specific nature of the texts. A comparison to the most common words found in texts of the “Confessional Materials” category (Table 1.2), in fact, shows significant similarities, not only in the specific words on the list but also the relative frequency of those words.

The lexical variation between these two groups of texts is obscured by these very high-frequency function words. By removing even a small subset of those words from the overall frequency counts, the differences between these two groups of texts are much more apparent. We removed a total of twenty “stop words,” i.e., high-frequency

Table 1.3. Most frequent words in the *DOE* “Prayer” category with stop words removed.

Rank	Word	Relative Frequency
1	Ic	.037
2	Me	.033
3	þu	.030
4	drithen	.020
5	Min	.019
6	Forgif	.019
7	Ealra	.008
8	Ne	.008
9	Synna	.008
10	þinum	.008

Table 1.4. Most frequent words in the *DOE* “Confessional Materials” section with stop words removed.

Rank	Word	Relative Frequency
1	He	.028
2	His	.023
3	Man	.020
4	Gif	.018
5	Ne	.018
6	Fæste	.016
7	Se	.015
8	Swa	.014
9	Hit	.012
10	Him	.011

words that might otherwise skew analysis, from both sets of texts and ran the frequency counts again.¹⁹ The results, seen in [Tables 1.3](#) and [1.4](#), demonstrate a much clearer connection between the highest-frequency words in the categories and the genres of the texts contained therein.

¹⁹ The stop words in question are: *ac, and, for, ge, in, mid, of, on, oþþe, to, þa, þam, þara, þæm, þæs, þæt, þære, þe, þurh, and wiþ*.

Table 1.5. Most frequent words in the four “Confessional Prayers” (B11.9.1, B11.9.3.1, B11.9.3.2, and B11.9.4).

Rank	Word	Relative Frequency
1	Ic	.059
2	Me	.026
3	Drihten	.018
4	þu	.018
5	God	.016
6	Min	.016
7	Ne	.013
8	Eallum	.011
9	Minum	.011
10	Eom	.010

The five stop words (*and*, *þe*, *on*, *þæt*, and *to*) on the original list for the prayer category were replaced largely by words that either reinforce the personal form of prayer (specifically, *þinum*) or highlight aspects of the content of the texts (*forgif* and *synna*, specifically). Similar changes are seen in the revised count of the confessional category, where the third-person form of the texts is emphasized by the appearance of the list of *hit* and *him*, and the penitential content of the category is represented by the appearance of *fæste*.

Perhaps more significantly, the two revised lists are very different from each other. With the stop words removed, only one word appears on both lists (and that word, *ne*, could easily have been included in the list of stop words, based on its almost purely grammatical function). The differences in the two lists—the types of pronouns, the presence or absence of direct references to God, etc.—are a product of the differences between personal prayers and confessional and penitential manuals. When a similar count is performed for the four “confessional prayers” that are included by the DOE in category B11, the similarities to the list of common words in the prayer group is immediately apparent.

Six of the words on the two lists (including the top four items on both) are exactly the same, and of the four words in Table 1.5 that do not appear in Table 1.3, three of them are simply different inflected forms of words that do appear in the other list. In other words, based simply on a count of the ten most frequent words (excluding stop words), the four contested “confessional prayers” clearly have more in common with other Old English prayers than with the confessional and penitential texts with which they have been categorized by the DOE.

A list of ten words is a small piece of data, however, especially since those words have been taken not from an individual text but from relatively large groups of texts. The

possibility of significant variations between texts inside that group—perhaps a single Old English prayer uses the word *forþif* a disproportionately large number of times, for example—raises questions about the integrity of the findings. What is needed is a way to look not at individual words in isolation but rather at the patterns of word frequency throughout these large groups. Using a statistical approach such as cluster analysis to analyze the lexical differences between the categories of texts more comprehensively helps to minimize the effect of those possible anomalies and provides a much clearer picture of the relationships among the texts being studied.

Cluster analysis is a basic data analysis technique designed to group objects in terms of their similarities. As seen, for example, in the lexomics model, this method consists of breaking a text into chunks of a predetermined size and using word frequency counts to calculate a mathematical distance between each chunk and the other chunks in the text. A relative frequency for each word in the chunk (i.e., the number of appearances of that word divided by the total number of words in the chunk) is determined, and then those relative frequencies are compared between each chunk and all other chunks. Chunks with more similar word frequency distribution are calculated to be closer in distance and, thus, more closely related than chunks with more variant distribution patterns. As mentioned above, the lexomics approach has tended to focus on individual texts, but a similar methodology can be used to compare groups of texts.

As a test case, we compiled two “sampler” datasets of Old English prose and poetry, taken more or less arbitrarily from the *DOE* Corpus of Old English.²⁰ As with the prayer and confessional materials datasets described above, each of these two datasets was scrubbed of punctuation, capitalization, and other extralexical information and then broken into chunks of 1,000 words each.²¹ Because, as mentioned above, we had determined that some subset of frequent words is more useful than overall word frequency counts for examining large groups of texts, we then identified the most common words in each chunk, here defined as words that appear at least fifteen times. This process yielded two lists of representative words that we call “fingerprints”: one for the poetry dataset and one for the prose. By combining these two lists and then removing any words that appeared on both, we were left with a master list of ninety-eight words. Using the hierarchical cluster analysis tool available in the software package R,²² we then compared the frequency of all ninety-eight of these words in each of the chunks of the

20 The texts in the poetry dataset were: “The Wanderer,” “Gifts of Men,” “Precepts,” “The Seafarer,” “Vainglory,” “Widsith,” “The Fortunes of Men,” “The Battle of Maldon,” “The Battle of Brunanburh,” “A Summons to Prayer,” “The Lord’s Prayer II,” “The Gloria I,” “The Lord’s Prayer III,” “The Creed,” and an excerpt (lines 828–1421) of *Beowulf*. The prose texts included one homily apiece from Aelfric and Wulfstan, three anonymous saints’ lives, and excerpts from the Old English versions of the *Benedictine Rule* and Bede’s *Ecclesiastical History* and the Alfredian translation of Orosius. The choice of texts was mostly random, though some effort was made to ensure representation of texts from multiple authors and time periods.

21 The size of the chunks is mostly arbitrary, though the size of the sample being studied obviously plays a role in that determination. See Drout et al., “Of Dendrogrammatology,” 313–14.

22 R Core Team, *R: A Language and Environment for Statistical Computing*.

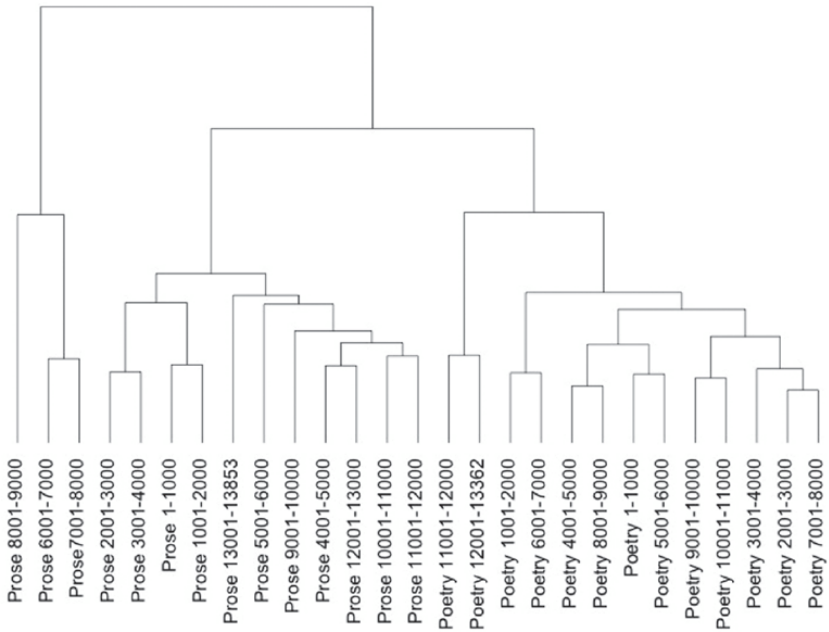


Figure 1.1. Dendrogram showing fingerprint word distribution in Old English poetry and prose “sampler” files.

two datasets.²³ Cluster analysis results are presented in branching diagrams known as dendrograms; the dendrogram created when we compared these two sampler files is shown in [Figure 1.1](#).

In this dendrogram, each chunk of the original files is labelled as either poetry or prose and with the word boundaries of the original file that the chunk contains; so “Poetry 1–1000” contains the first 1,000 words of the poetry sampler file, and so on. It is important to note here that the chunk boundaries are not related to the boundaries between individual texts inside the chunks. The first chunk of the poetry file contained all of “The Wanderer” and parts of “Gifts of Men,” for example, while the second chunk contained the remaining words from “Gifts of Men,” all of “Precepts,” and parts of “The Seafarer.” On the dendrogram, the chunks are grouped according to the distance between

23 As in the lexomic model, we use the Euclidean metric to calculate the distance between any two chunks. Our clustering model is slightly different, however. Lexomics uses what is known as the average linking method to determine the distance between chunks, while we use an approach known as Ward’s method, which calculates distance based on a “center of gravity” in each chunk. The two linking methods yielded similar results in our tests, but Ward’s method is designed to reduce the amount of “noise” in each cluster and, thus, appears to be more useful in comparing large groups of texts that may have only some elements in common. See Husson, Lê, and Pagès, *Exploratory Multivariate Analysis by Example Using R*.

them when frequency of the fingerprint words inside each chunk is compared. So on the right hand side of the dendrogram, we can see that the chunk containing words 2001–3000 of the poetry file are very similar to the chunk containing words 7001–8000 of the same file. The chunk labelled Poetry 3001–4000 is not quite as similar, but that chunk is still more closely connected to those two chunks than any other chunk in the dendrogram. The height of the stems connecting the chunks represents the relative similarity between the chunks, so while the chunk labelled Prose 8001–9000 (on the far left-hand side of the dendrogram) is shown to have some similarities to Prose 6001–7000 and Prose 7001–8000, the connection there is much less close than that between Poetry 3001–4000 (on the far right-hand side) and the two chunks closest to it.

Several interesting findings are suggested in this dendrogram, but clearly the most important one for our purposes is the absolute separation between the poetry and prose chunks. Even though the individual texts that comprised the sampler files were chosen arbitrarily—and even though those texts varied in authorship, date, and individual style—the cluster analysis algorithm was able to recognize the difference between Old English prose and poetry with complete reliability, based only on a comparison of a relatively small set of frequently appearing words. Interestingly, the specific words in the fingerprint do not seem to be directly related to the genre of the texts in the sampler files. Only one of the words (*folde*) in the poetry fingerprint list is a recognized poetic word. The difference identified by the cluster analysis in this case, in other words, is mainly a difference in the distribution of common Old English words, rather than the result of a genre-specific lexicon.

Having established the efficacy of this method in detecting genre differences, we applied the same approach to the texts categorized by the *DOE* as prayers and confessional materials. As we had when calculating overall word frequency, we compiled all of the texts in the *DOE* “Prayer” category into a single file, and all of the texts from the “Confessional Materials” category, again with the exception of the four contested confessional prayers, into a second file. After removing stop words, we separated both files into chunks of 1,000 words each and calculated overall word frequency for the two sets. To create the fingerprint lists, we identified the words that appear at least twenty times in each set and then discarded the words that appeared on both lists. We were left with a master list of 146 fingerprint words, 130 specific to the confessional group and 16 specific to the prayer group.²⁴ Comparing the two groups of texts according to the combined fingerprint list resulted in the dendrogram seen in [Figure 1.2](#):

Several points are evident in this dendrogram. Variations within each of the groups can be seen, for example, as can the fact that slightly more variation exists within the small corpus of Old English prayers than in the much larger corpus of confessional materials. But the most notable finding here is clearly the overarching separation between the two groups. The variation within the two groups is dwarfed by the variation between them,

24 The imbalance of these lists is partially a product of the difference in size of the two groups. The confessional material category (B11) is almost five times the size of the prayer category.

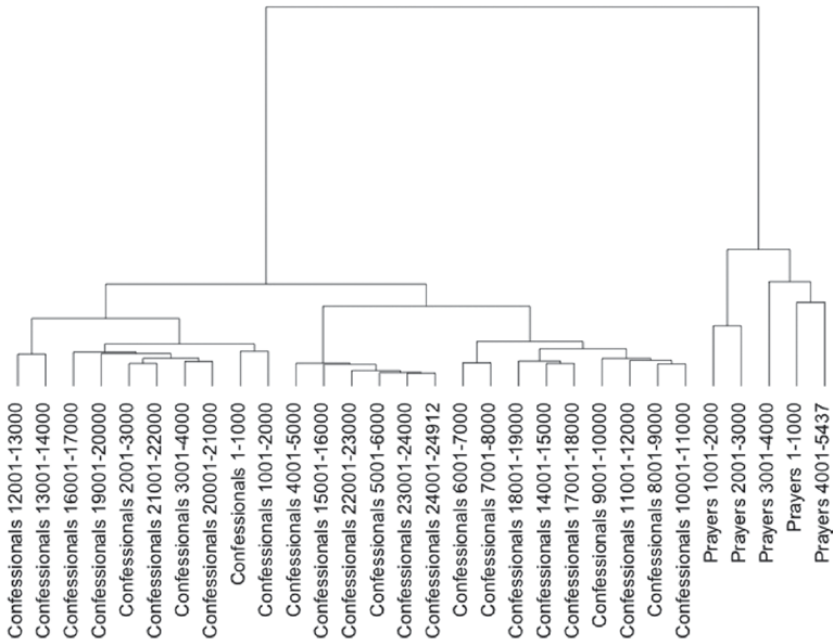


Figure 1.2. Prayers and confessional materials using a combined 146-word fingerprint.

as seen in the height of the stems connecting the two main clades.²⁵ Based purely on the distribution of the fingerprint words we identified, the boundary between Old English prayers and Old English confessional materials is both pronounced and absolute.

The same separation is seen if we define the fingerprint set according to relative frequency instead of absolute frequency. We ran the same analysis using the thirty most frequent words in the two sets as the fingerprint, as seen in [Figure 1.3](#). Some of the specific groupings inside the category of confessional materials shifted slightly as a result of this change, but the overall separation between the two groups is unaffected. In each of the two sets, the list of thirty most frequent words represents approximately 30% of the total word distribution in the categories, a portion large enough, in other words, to allow extrapolation to larger patterns of word use throughout the group. Remarkably, however, we see a virtually identical separation even when using only the ten most frequent words in the two groups, a fact which demonstrates two important attributes of these datasets. As seen in [Tables 1.3](#) and [1.4](#), the ten most frequent words in these groups still account for between 15% and 20% of the total word distribution in the two categories. And as the cluster analysis demonstrates, the distribution of those words is

²⁵ In cluster analysis, a *clade* is simply a grouping of at least two chunks in a dendrogram. See Drout et al., “Of Dendrogrammatology,” 306.

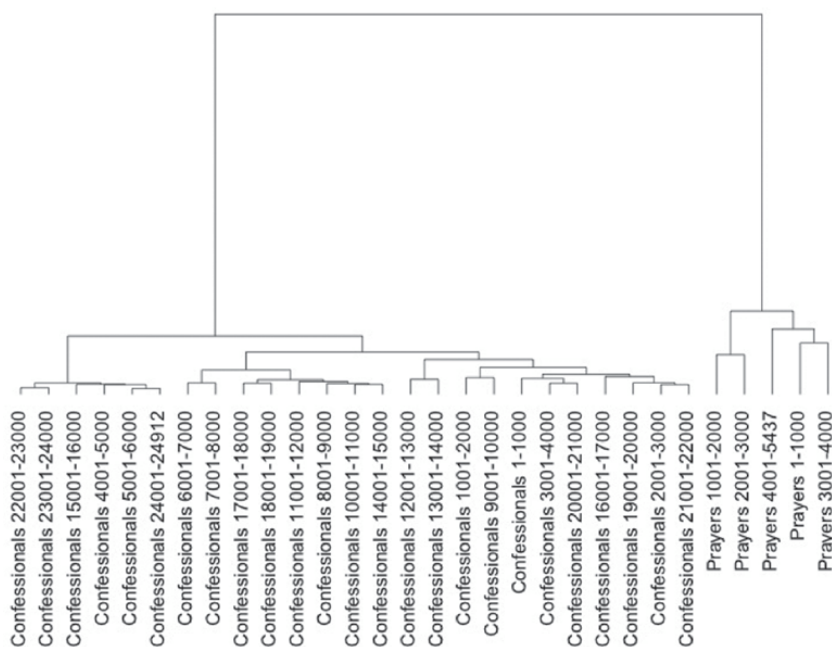


Figure 1.3. Prayers and confessional materials using the thirty most frequent words in each set.

relatively consistent across the individual texts inside those two categories, a characteristic that speaks to the relatively uniform style of both Old English prayers and penitential manuals. Broader genres that allow more latitude regarding content and style would likely require a somewhat larger fingerprint word list in order to accurately determine such boundaries. When we ran a similar ten-word fingerprint analysis on the poetry and prose sampler files, for example, the result was a much more scattered dendrogram that did not identify the genre boundary precisely.

Having suitably established the separation between prayers and confessional materials, we then included the four contested “Confessional Prayers” into the analysis. Using our original 146-word combined fingerprint, we generated the dendrogram seen in [Figure 1.4](#). Because the four additional texts consist of only a little more than 2,000 words in total, they are represented by just two chunks on the dendrogram, here labelled as Confessional Prayers 1–1000 and 1001–2320. Despite the small sample size, however, those two chunks are unambiguously grouped with the other Old English prayer chunks, suggesting that the devotional nature of the four texts is more strongly represented by their word use patterns than are any confessional elements found therein. Including those four confessional prayers in the analysis does affect the separation between the two larger groups, however. Although the texts in the prayer category are still a distinct group in [Figure 1.3](#), the dendrogram does show some chunks from the confessional category to be more similar to the combined “prayer/confessional prayer” group than to the

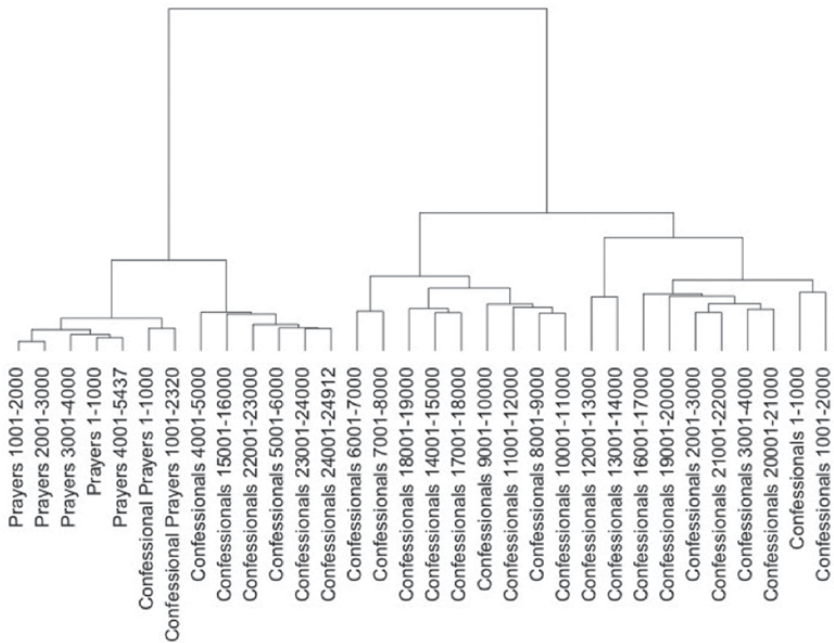


Figure 1.4. Prayers, confessional materials, and the four “Confessional Prayers,” using a combined 146-word fingerprint.

other chunks from the confessional materials. Apparently the four confessional prayers have a kind of magnetic effect on those chunks, sufficiently varying the clade in which they are joined to the other prayers to make them somewhat similar to a portion of the texts in B11.

When we switched to the fingerprint based on the thirty most frequent words in the two categories, the results were slightly less clear. As seen in [Figure 1.5](#), based on this fingerprint set, the four contested prayers are still grouped closely with the other Old English prayers, but now the main group of prayers is seen as slightly more similar to those six outlying chunks from the confessional category than they are to the four confessional prayers. In other words, though this analysis still found meaningful similarities between the four contested texts and the body of Old English prayer in general, it was less successful at identifying a clear boundary between genres.

Surprisingly, when we reduced the fingerprint set to the ten most frequent words, the groupings resolved in a dendrogram very similar to that seen in [Figure 1.4](#). The new dendrogram, seen in [Figure 1.6](#), again groups the four contested prayers with the other Old English prayers and clearly separates the prayers from the confessional materials. This finding shows the importance of determining the most effective fingerprint list for the specific groups of texts being analyzed. Though all three of the dendrograms agreed that the four contested texts are generally more similar to the corpus of Old English prayers than to the rest of the confessional category in which they have been placed by

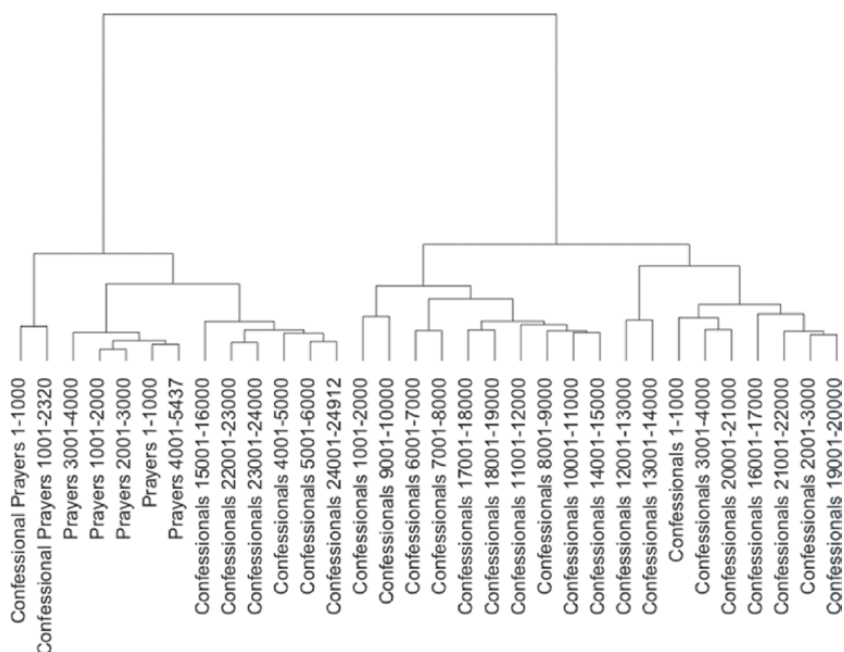


Figure 1.5. Prayers, confessional materials, and the four “Confessional Prayers,” using the thirty most frequent words.

the *DOE*, the specific degree of similarity detected and, importantly, the precision of the genre boundary marker between the prayers and the confessionals varied somewhat based on the specific set of words counted.

This variation should not be seen as a limitation of the fingerprint method. Rather, it demonstrates the sensitivity of the process and the vital role of the researcher. Just as the process requires multiple iterations to find the most effective set of fingerprint words, the judgment of the researcher is also required for identifying the best list of stop words and determining how much regularization of the source texts is required. In other words, the role of the researcher in this process is not a passive one. Specifically, the nature of the groups being examined needs to be carefully considered. Because prayers and confessionals are two very closely related genres of Old English prose, much of their vocabulary is shared in common. In this case, the fingerprint method was successful in multiple iterations, whether focusing on a very small list of distinctive words (i.e., the ten most frequent words in both datasets) or a significantly larger list (consisting of all words that appear at least twenty times). The fact that the results varied slightly when using a fingerprint list of moderate size (i.e., the thirty most frequent words in each set) only speaks to the similarities between these genres.

Ultimately, the method was successful in each attempt, in that it consistently recognized the distinctions between the two genres and correctly identified the four

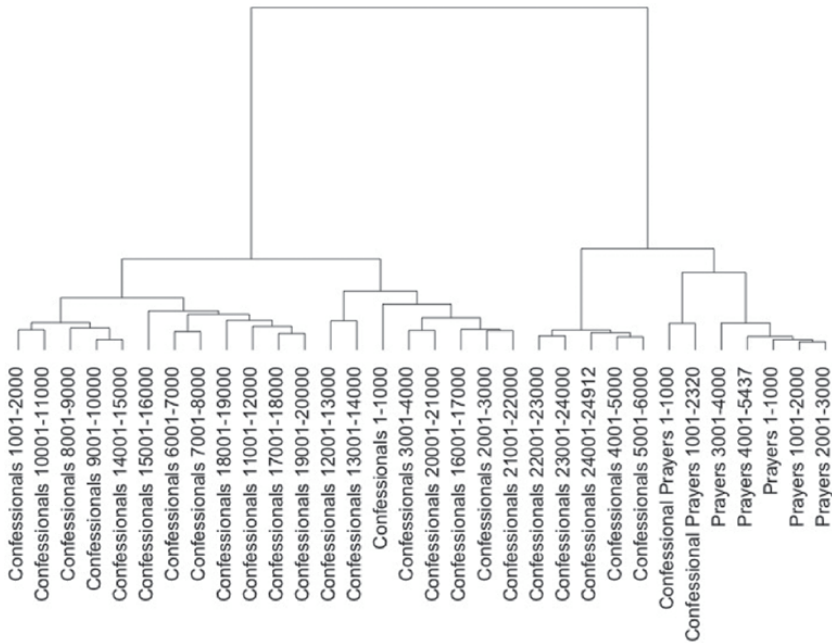


Figure 1.6. Prayers, confessional materials, and the four “Confessional Prayers,” using the ten most frequent words.

disputed texts as prayers, rather than confessional materials. The method has great potential in enhancing our understanding of the way we categorize Old English texts. It may contribute to the critical debate regarding Aelfric’s “rhythmical prose,” for example. Thomas Bredehoft has convincingly argued for classifying Aelfric’s *Lives of Saints* as verse, based purely on metrical criteria.²⁶ Since, as we have seen, Old English poetry and prose tend to contain markedly different word use patterns, it would be interesting to compare representative texts from the *Lives of Saints* to an established fingerprint of Old English verse. Perhaps in such a case, as with the case of Old English prayers, statistical analysis will provide us with evidence not immediately visible using traditional methods. Ultimately, the computer may allow us an otherwise unavailable glimpse into the way Anglo-Saxon authors and readers thought about and used the idea of textual genre.

²⁶ Bredehoft, *Early English Metre*. In a later monograph, Bredehoft expands his argument to treat the appearance of small lexical units in Aelfric’s later works as poetic formulae. See Bredehoft, *Authors, Audiences, and Old English Verse*, 146–70. Interestingly, in an appendix to the later book (208–16), Bredehoft argues that two texts classified by the *DOE* as prayers (B12.4.3.1 and B12.4.7) should also be considered verse compositions.

Bibliography

- Allison, Sarah, Ryan Heuser, Matthew Jockers, Franco Moretti, and Michael Witmore. "Quantitative Formalism: An Experiment." *N+1* 13 (2012): 81–108.
- Bately, Janet. "Did King Alfred Actually Translate Anything? The Integrity of the Alfredian Canon Revisited." *Medium Aevum* 78 (2009): 189–215.
- Battles, Paul. "Toward a Theory of Old English Poetic Genres: Epic, Elegy, Wisdom Poetry, and the 'Traditional Opening.'" *Studies in Philology* 111 (2014): 1–33.
- Bredenhof, Thomas. *Authors, Audiences, and Old English Verse*. Toronto: University of Toronto Press, 2009.
- . *Early English Metre*. Toronto: University of Toronto Press, 2005.
- Burrows, J. F. "Questions of Authorship: Attribution and Beyond." *Computers in the Humanities* 37 (2003): 5–32.
- Craig, Hugh. "Stylistic Analysis and Authorship Studies." In *A Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens, and John Unsworth, 273–88. Oxford: Blackwell, 2004.
- Drout, Michael D. C. *Tradition and Influence in Anglo-Saxon Literature: An Evolutionary, Cognitivist Approach*. New York: Palgrave Macmillan, 2013.
- Drout, Michael D. C., Michael J. Kahn, Mark D. LeBlanc, and Christina Nelson. "Of Dendrogrammatology: Lexomic Methods for Analyzing Relationships among Old English Poems." *Journal of English and Germanic Philology* 110 (2011): 301–36.
- Frank, Roberta, and Angus Cameron. *A Plan for the Dictionary of Old English*. Toronto: University of Toronto Press, 1973.
- Gill, Paramjit S., Tim B. Swartz, and Michael Treschow. "A Stylometric Analysis of King Alfred's Literary Works." *Journal of Applied Statistics* 34 (2007): 1251–58.
- Hoover, David L. "Testing Burrows's Delta." *Literary and Linguistic Computing* 19 (2004): 453–75.
- Husson, François, Sébastien Lê, and Jérôme Pagès. *Exploratory Multivariate Analysis by Example Using R*. Boca Raton: CRC Press, 2011.
- Ker, N. R. *Catalogue of Manuscripts Containing Anglo-Saxon*. Oxford: Clarendon, 1957.
- Logeman, Henri. "Anglo-Saxonica Minora." *Anglia* 11 (1889): 97–120.
- Pulsiano, Philip, and John McGowan. "Four Unedited Prayers in London, British Library, Cotton Tiberius A. iii." *Mediaeval Studies* 56 (1994): 189–216.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, 2015. www.R-project.org/
- Rushforth, Rebecca. "The Barrow Knight, The Bristol Bibliographer and a Lost Old English Prayer." *Transactions of the Cambridge Bibliographical Society* 12 (2003): 372–92.
- Smith, William H. "The Tradition of Vernacular Prayer in Anglo-Saxon England" PhD dissertation, University of North Carolina at Chapel Hill, 2005.

Chapter 2

IF (NOT “QUANTIZE, CLICK, AND CONCLUDE”) {DIGITAL METHODS IN MEDIEVAL STUDIES}¹

KATAYOUN TORABI

MEDIEVALISTS WHO WISH to do research with digital tools may find their options limited because of the incipient nature of the practice. Technological limitations, specifically the difficulty of digitally recognizing and rendering medieval texts due to the limited availability of tools that can accommodate special characters,² create constraints, in turn, on discipline development. Another constraint lies in the paucity of work that documents digital methodologies in medieval scholarship, leaving the would-be digital medievalist very little to draw upon in terms of established good practices. This has led to a kind of methodological misunderstanding on the part of certain humanist scholars whose reliance upon digital tools leaves their work open to the critique that because computer-assisted analysis can be reductive, undue authority is ascribed to their findings.

I insist here that whenever digital tools are utilized in medieval studies, the methodologies with which they are applied should be paired with and grounded in traditional humanist analysis—that is, a close reading of the text that carefully scrutinizes information produced by computer-aided analysis. In this chapter, I discuss two recent projects that I have worked on that use web-based digital tools to analyze Old and Middle English texts. I will demonstrate that computer-assisted textual analysis works best when paired with traditional humanist analysis, with the former supplementing rather than replacing the latter. I also demonstrate that such projects can be open-ended, and argue that digital tools are not meant to be end-to-end problem-solving technologies, but can be effectively used to affirm or deny hypotheses, assist in the interpretation of texts, and help discover significant patterns that are undetectable by the unaided eye in order to generate new questions for further investigation using traditional methods.

Finding the correct digital research tool, however, can be a lengthy process of trial and error. The Text Analysis Portal for Research (TAPoR)³ and the Digital Research

¹ The title is written in what software developers refer to as *pseudo-code*. It borrows syntactical conventions mainly from Java, C, and Python. The title encapsulates this chapter’s central argument, which acknowledges that computer-assisted analysis can be reductive, but contends that the methodologies with which digital tools are applied should be grounded in traditional humanist analysis.

² Many digital research tools for scholarly use are designed for modern English texts and may be unable to process special characters—such as *thorn* (þ), *eth* (ð), and *ash* (æ)—and many are unable to accommodate morphological and orthographical variation inherent in Old and Middle English texts.

³ Rockwell, Uszkalo, and Sinclair, The Text Analysis Portal for Research (TAPoR).

Tools Directory (DiRT)⁴ are good places to start for humanists interested in doing digital research because they provide access to various commercial and open-source digital research tools. DiRT's search options allow users to select digital tools according to research objectives, cost, and platform; while TAPoR provides a curated list of tools organized according to usage, ease of use, and research objectives, providing users with the option of excluding tools that are complex and difficult, time consuming, and of limited use. While both sites allow users to exclude tools that are still in their beta phase, it is important to note that even well-established and widely used tools are typically revised, updated, and improved. A tool that may not initially be a good fit might work better after improvements over time. As such, the best way to approach the incorporation of digital tools into medieval research is by experimenting and being open to the possibility of failure. Failure can occur when a tool turns out to be a poor match for the project. It can also occur when the tool works, but does not produce the looked-for results to support a research claim. Regardless, such failure is an invaluable part of the process that should be documented for future research.

The projects I discuss in this chapter involve using digital tools to analyze Old and Middle English texts. The first project compares passages in *Beowulf* and *Blickling Homily XVII* using Lexomics, an online web-based tool suite developed by a group at Wheaton College.⁵ The second project uses Lexomics and another web-based tool, Voyant,⁶ to compare sections of text and track word usage in the *Canterbury Tales*. These projects demonstrate the need for experimentation with tools in order to determine what does and does not work, the open-endedness of projects, the need for documentation throughout the process in order to maintain transparency for the audience and for possible future collaboration with other scholars, and most significantly, the importance of anchoring all computer-aided research in traditional humanist analysis.

Old English: *Beowulf* and *Blickling Homily XVII* Project

Using a cluster analysis tool developed by the Lexomics team at Wheaton College,⁷ I compared the description of St. Paul's vision of hell in *Blickling Homily XVII*⁸ and the passages in *Beowulf* that describe Grendel's mere (lines 1357b–1382b and 1405–1417).⁹ Because *Blickling Homily XVII*'s description of hell as a dark place haunted by frost-covered trees,

4 Spiro, Digital Research Tools Directory (DiRT).

5 Drout, Kahn, LeBlanc, and Nelson, "Of Dendrogrammatology: Lexomics Methods for Analyzing the Relationships among Old English Poems," 301–36. The Lexomics tool suite may be accessed at: <http://wheatoncollege.edu/lexomics>.

6 Sinclair, Rockwell, and the Voyant Tools Team, Voyant Tools.

7 Kleinman, LeBlanc, and Drout, "Lexomics Tools."

8 *Blickling Homily XVII* is part of a collection of anonymous homilies composed in Old English prose. See Scragg, "Corpus," 223–77.

9 This project was developed into an article by Katayoun Torabi, "Two New Approaches to Exploring Monstrous Landscapes in *Beowulf* and *Blickling Homily XVII*."

wolves, and water monsters is remarkably similar to the description of Grendel's icy, monster-infested abode in *Beowulf*, scholars have long discussed the possibility of one text borrowing from the other.¹⁰ My goal for this project was to use a digital tool that performs cluster analysis in order to examine similarities in the frequency and distribution of words in the monster-mere passage in *Beowulf* and St. Paul's vision of hell in *Blickling Homily XVII* that would suggest a close textual relationship between those sections. I hypothesized that the many lexical and phraseological similarities between these passages would not only be visible to cluster analysis, but that the lexomic method would also detect a distinctive lexical pattern in *Beowulf*'s monster-mere passage, which would suggest an outside source. Although the sort of analysis that can be performed by digital tools can offer certain insights not readily available via traditional means, it can never replace a careful close reading of the text. As such, I followed the lexomic "reading" of the poem with a more traditional formula density analysis, which mapped formulaic language in and around the passage describing Grendel's mere in *Beowulf*. I posited that a difference in the number of formulaic words or phrases between the monster-mere passage and the surrounding text would suggest an outside source for that passage. Both the lexomic method and the formula density analysis, however, disproved my original hypothesis and led me to conclude that similar word frequency and distribution and shared formulaic words and phrases were not necessarily suggestive of any borrowing by either the *Beowulf* poet or *Blickling* homilist; but rather, any similarities between the passages likely resulted from both writers using a shared set of cultural conventions for describing monstrous landscapes in the Anglo-Saxon tradition when composing their work.

The lexomic method, which involves a four-step process, calculates word frequency and distribution in a text, and visualizes lexical similarities and differences in target text(s) as a branching tree diagram, called a dendrogram.¹¹ The user begins by putting the text he or she wishes to analyze in a text editor, such as Notepad++, and then dividing that text into sections or chunks of equal word count, using the Lexos cutter. In [Figure 2.1](#) below, for example, *Beowulf* and *Blickling Homily XVII* are divided into forty separate chunks, each containing 500 words. Chunks one through thirty-five, on the left side of the dendrogram are from *Beowulf* (labelled "B") and chunks one through five on the far right (labelled "BH") are from *Blickling Homily XVII*. The chunked texts are then processed by the Lexos clustering tool, which

calculates each word's relative frequency—that is, the number of times a word appears in proportion to the entire number of words in a text. The relative similarity or dissimilarity of the text chunks are then calculated using a Euclidean distance formula, and hierarchical agglomerative clustering is employed to

10 The connection between *Blickling XVII* and *Beowulf* was first noted by Morris in his edition of the *Blickling Homilies*, v–viii. It was followed by Brown, "Beowulf and the *Blickling Homilies*," 905–16; and Malone, "Grendel and His Abode," 297–308. The debate about a possible connection between the two texts is discussed in further detail in Torabi, "Two New Approaches," 165–67.

11 A description of the lexomic method is detailed in Torabi, "Two New Approaches," 167.

construct a dendrogram—a branching tree diagram—that provides a graphical visualization of the data, assigning individual sections of a text to a clade or branch. The clustering of the branches and their length visually represent degrees of similarity between text chunks.¹²

Similar chunks of text will cluster closer together in the tree diagram and have shorter connecting branches; and dissimilar chunks will cluster farther apart and have longer branches.¹³

In [Figure 2.1](#), the *Blickling Homily XVII* chunks clustered apart on long vertical branches away from the “B” branches, indicating that not only is *Blickling Homily XVII* lexically distinct from *Beowulf*, but the monster-mere passage, in chunk “B_15,” and St. Paul’s vision of hell, in chunk “BH_5,” are also quite distinct from one another as they clustered at a significant distance. Furthermore, chunk “B_15” failed to cluster apart from the rest of *Beowulf*, which suggests that the monster-mere passage does not have a source that is distinct from the rest of the text.¹⁴ The results of the cluster analysis did not support my original hypothesis and produced no evidence that would suggest a textual relationship between *Beowulf*’s monster-mere passage and St. Paul’s vision of hell in *Blickling Homily XVII*. Even when chunk sizes were increased to 1,000 words, the *Blickling Homily XVII* and *Beowulf* texts continued to cluster at a significant distance from one another. Lexomic analysis did not provide any evidence that the sections describing hellish landscapes in *Blickling Homily XVII* and *Beowulf* have any textual connection, as they consistently clustered at a great distance from one another. At this point in the project additional, more precise tests were needed in order to determine if the clustering pattern had occurred because there really is no textual relationship between the *Beowulf* and *Blickling Homily XVII* passages or if the tools I used for this project presented an as-yet-unidentified technological constraint.

Running more tests, however, would require a considerable time commitment. When the developers of Lexomics first rolled out their tool suite in 2009 it did not include many of the time-saving text-processing tools that are now readily available on the Lexomics website. For instance, because I was working from an OCR’d edition of Frederick Klaeber’s *Beowulf and the Fight at Finnsburg*,¹⁵ it would have been necessary to manually convert all words in the target texts to lower case and remove punctuation,

¹² Torabi, “Two New Approaches,” 167.

¹³ A full explanation of how hierarchical agglomerative cluster analysis and the Euclidean distance metric function are used in the creation of dendrograms can be found in the appendix of Drout et al., “Of Dendrogrammatology,” 336.

¹⁴ Lexomic cluster analysis successfully detects lexical boundaries within and among various texts, including *Azarias*, *Daniel*, and *Genesis A and B*. The Wheaton Group’s experiments were published in Drout et al., “Of Dendrogrammatology,” and more recently in Drout’s *Tradition and Influence*. Their findings are also referenced in note 12 in Torabi, “Two New Approaches,” 178.

¹⁵ Klaeber, ed., *Beowulf and the Fight at Finnsburg*, 4th ed. Because Klaeber’s edition of *Beowulf* was not available online. I used an optical character recognition reader to convert images of Klaeber’s edition into PDF format, with machine-encoded text. The editors of the 4th edition meticulously documented all of their modifications, which allowed me to take into account “any and

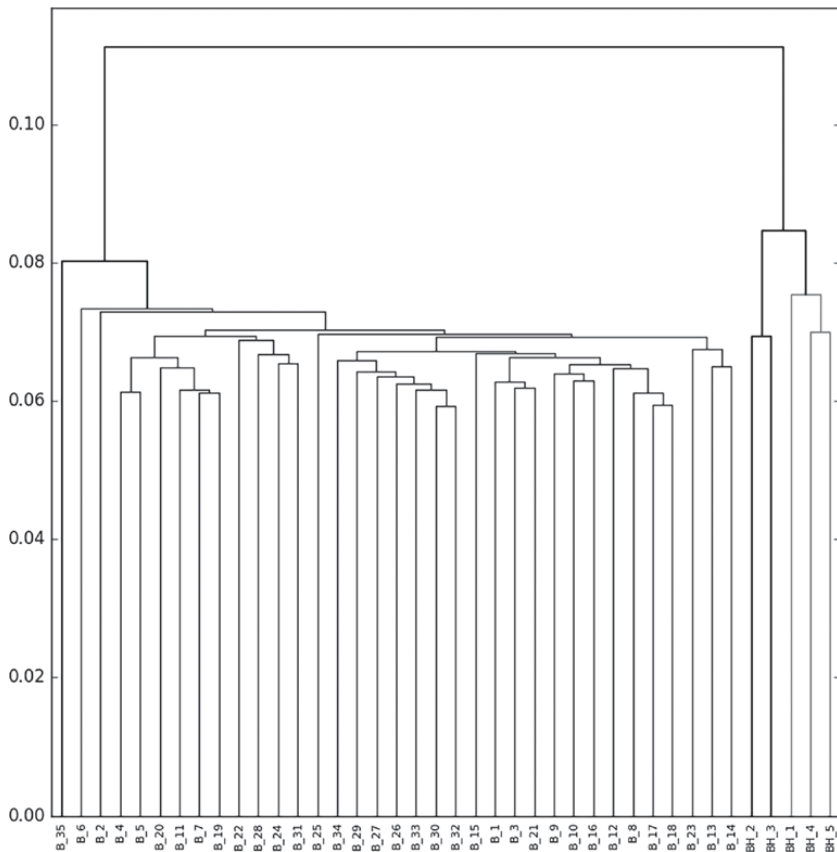


Figure 2.1. Combined *Beowulf* and *Blickling Homily XVII* 500-word chunks (Lexomics Tools).¹⁶

numbers, and function words (pronouns, demonstratives, articles, conjunctions, causal words, certain adverbs, and prepositions), before normalizing special characters and lemmatizing lexemes (converting words to their uninflected and unconjugated forms). Lemmatized and non-lemmatized versions of *Beowulf* and *Blickling Homily XVII* would allow me to see how the presence or absence of inflectional endings and function words affected lexomic cluster analysis. Unfortunately, the preparation of such texts for cluster analysis would have been such a labour-intensive process that I decided to step away

all rewodings, additions, and deletions made by the editors and to take into account all editorial changes when performing lexomic analysis.” See note 13 in Torabi, “Two New Approaches,” 178 for more information about my selection of Klaeber’s edition of *Beowulf* for this project.

16 Kleinman, LeBlanc and Drout, “Lexomics Tools.” All Dendrograms have been generated using Lexomics tools.

from the project until updated digital tools became available that would allow for faster and more efficient processing of texts for cluster analysis.

At this point, I came to realize (and reluctantly accepted) that digital projects are sometimes necessarily open-ended. As new digital text analysis tools are created or updated, setting aside a project that utilizes digital tools as part of the research process is an acceptable and sometimes necessary option until technology becomes available to address certain research questions or problems. This is especially true for humanists who do not have the time nor the inclination to learn computer coding in order to create their own set of digital tools. The limitations of the technology, therefore, are paramount in the use of digital tools as part of the research process. Before I set aside the *Beowulf-Blickling* project, however, I carefully documented each step of the process in order to facilitate a less complicated return to the work. Such documentation would also allow others to pick up where I left off if my research interests did not return to this problem. It should also be noted here that whenever texts are prepared for processing by digital tools, changes are made to and categories are imposed on the text that are artificial and highly subjective. For example, normalizing a text by eliminating grammatical and orthographical variants caused by dialectal variation arbitrarily privileges one variation of a text over another. This is also true for the process of lemmatization. What is important, therefore, is that whatever criteria are used for the preparation of a text are applied consistently throughout a single study and that the decision-making process is clearly documented. Doing so not only eases the transition back into a project after a hiatus, but appropriately enables scholarly collaboration.

When I revisited the project a year and a half later, in early 2013, new tools and updates had become available on the Lexomics website, making cluster analysis a much faster, easier, and more efficient process. I was able to generate several dendrograms in a matter of minutes, which allowed me to analyze the texts as thoroughly as possible in order to complete the project. A new addition to the Lexomics tool suite was the Lexos text scrubber, which significantly reduced the amount of time required to produce dendrograms by allowing me to quickly lemmatize texts, remove function words, upload a list of character replacements, and upload instructions for handling non-Unicode characters. The addition of the Lexos text scrubber most significantly impacted my project by drastically reducing the effort it took to prepare the data for cluster analysis.

As a result of the increases in efficiency, I was able to produce over thirty combined 500- and 1,000-word chunk dendrograms for the *Beowulf* and *Blickling Homily XVII* texts in under an hour. I generated dendrograms that excluded Latin words, function words, and others that excluded content words (nouns, adjectives, verbs, and most adverbs). I also generated dendrograms of lemmatized and non-lemmatized *Beowulf* and *Blickling Homily XVII* texts. In all instances, however, cluster analysis did not reveal any evidence that the section describing hell in *Blickling Homily XVII* has a textual relationship with the passages describing Grendel's mere in *Beowulf*, as they continued to cluster at a great distance from one another. These findings contradicted—or at least failed to support—what many scholars have claimed about the existence of a close relationship between *Beowulf* and *Blickling Homily XVII*.

I posited that the sections in *Blickling Homily XVII* and *Beowulf* describing monstrous landscapes of hell each consist of approximately 200 words and may be too small

to be accurately measured by the lexomic method. To test this hypothesis, I copied the main passage describing Grendel's mere (1357b–1382b) and inserted it into the Old English poem, *Genesis A* and *B*, at the very beginning of the text. *Genesis*, at 2,936 lines, is comparable in length with *Beowulf* and could, therefore, serve as a control to test whether the lexomic method is able to detect the presence of a short, alien passage that consists of twenty lines. After running a cluster analysis using eleven 1500-word chunks, the chunk containing the short *Beowulf* passage clustered closely with the rest of the *Genesis* text. Interestingly, cluster analysis correctly identified differences between *Genesis A* (lines 1–234 and 886–2,936) and *Genesis B* (lines 235–885), as the two clustered apart from one another—results that are consistent with the findings of the Lexomics group at Wheaton College.¹⁷ It appears, however, that the *Beowulf* passage that was only twenty lines long was too small to be detected, as the chunk containing it clustered with the rest of the text. After increasing the number of words in the *Beowulf* passage to 1,000 and appending it to the beginning of *Genesis*, a new cluster analysis revealed that the alien passage was indeed detected and the chunk clustered at a significant distance from the rest of *Genesis*. Based on this evidence, I concluded that although lexomic methodology accurately detected patterns within and among texts that had large enough sample sizes (such as *Guthlac A* and *B*, *Daniel*, and *Azarias*), it failed to measure patterns accurately in texts that have small test target sample sizes (less than 300 words).

Therefore, it is important to know when a certain tool is not the right fit for the problem or question being explored. When the Lexomics tool suite was first made available on the Wheaton College website, for the reasons stated above, I was unable to perform all of the necessary tests for my project and received results that I could not fully interpret at the time. Once the tool suite was expanded and existing tools were upgraded, however, I was able to very quickly generate several different kinds of dendrograms for the *Beowulf* and *Blickling Homily XVII* texts, which helped me to gain a better understanding of how and why my methods yielded certain results.¹⁸ The Lexomics tool suite is among the best for medieval scholarly research, as it is one of the few digital tool suites that can appropriately handle the various idiosyncrasies of Old and Middle English texts. New digital tools

¹⁷ Drout et al., “Of Dendrogrammatology,” and more recently in Drout’s *Tradition and Influence*.

¹⁸ It should be noted here that in April 2014 I presented the results of my lexomic cluster analysis in a paper titled, “‘Nis þæt heoru stow!’ A Lexomic Approach to the Monstrous Places in *Blickling Homily XVII* and *Beowulf*,” at the Presents of the Past Symposium at Texas A&M University. Michael Drout, who was in attendance, stated that the Lexomics team was adding a rolling window tool to their tool suite which would address the lexomic clustering method’s difficulty in comparing smaller sections of text. Drout, Kisor, Smith, Dennett, and Piirainen published *Beowulf Unlocked: New Evidence from Lexomic Analysis*, wherein they state that in a forthcoming publication titled *Beowulf Unlocked II: Evidence of Rolling Window Analysis* they will document their use of the new rolling window tool, which allows users to analyze segments of text that are too small for cluster analysis. Since the rolling window tool is now available on the Lexomics website, I may use it to take another look at the passages describing monstrous places in *Beowulf* and *Blickling Homily XVII*.

are always being developed and improved, however, and the most effective way to determine if they are compatible with medieval texts is to experiment. As mentioned above, the TAPoR and the DiRT websites are two excellent places to start. Many of the tools on these websites, however, are in their beta phase and rely on the bold and curious to use them and provide feedback to their authors. Choosing the correct tool can be a very time-consuming process that may deter some scholars from incorporating computational methods into humanities research. The best way to address this problem seems to be the production of more peer-reviewed articles documenting experiments with these tools.

In my case, although the target passages in *Beowulf* and *Blickling Homily XVII* were too small to be analyzed by the lexomic method at that time, my “failure” to produce the hoped-for results prompted me to ask another question: what if the text surrounding the relevant passages in each chunk was creating enough noise to skew results? I hypothesized that if the target passage (regardless of size) could be isolated and analyzed, along with its adjacent passages, then salient linguistic shifts may become visible to the lexomic method. I generated one last dendrogram—one that excluded all text in *Beowulf* except for the description of the main monster-mere passage (lines 1357b–1382b) and sixty lines that precede and follow it. The dendrogram, in this instance, revealed a slight separation of the monster-mere passage from its adjacent sections. In order to further investigate the significance of that separation, I shifted my attention back to the text and performed a close reading, which pointed to a higher than average presence of formulaic words and phrases in the monster-mere passage. The formula-rich language in the monster-mere passage, which was not present to the same degree in the adjacent text, may account for the separation. I concluded that although the presence of highly formulaic language in the monster-mere passage caused a significant linguistic shift that became visible to lexomic analysis, this evidence alone is not necessarily suggestive of an outside source for the monster-mere passage. Rather, the poet, in composing the passage describing the monstrous landscape, may have been drawing from a common, culturally embedded set of ideas and conventions for describing such landscapes.¹⁹

One significant problem encountered when I incorporated digital tools and methodologies into my research was writing about my work—especially writing the sections that detailed lexomic data analysis. I struggled with form and style as I attempted to explain the lexomic method, the various tests that I ran, and the results generated by those tests. This was a major stylistic departure from any piece of writing that I had produced in the past. I tried to write with clarity and to capture enough details without overburdening the readers with technical jargon that they would find tedious and irrelevant. What I produced was a method-driven narrative that discussed the experimental aspect of the project and focused on an analysis of Old English texts in a way that is legible within extant medieval discourse. In other words, combining digital and traditional methodologies also necessitates developing a new way of thinking through the narrative process.

My analysis of monstrous places in *Beowulf* and *Blickling Homily XVII* paired computer-based analysis with more traditional humanist analysis in order to offer new

¹⁹ For a more detailed explanation of how I used formula density analysis in this project, please refer to Torabi, “Two New Approaches,” 172–76.

insights into an ongoing debate about the possible connection between the texts in the field of Old English literature. Lexomic hierarchical agglomerative clustering produced dendrograms that visualized salient shifts in linguistic patterns in *Beowulf*, which then prompted me to perform a close reading in order to investigate the cause of those shifts. I chose to discuss that project in this chapter because it speaks to the anxieties critics of the digital humanities have about computational methods replacing more traditional forms of scholarship.²⁰ Certain critics, for example, contend that computer-assisted analysis is not only reductive, but assigns the work of thinking to a computer.²¹ Katherine Hayles argues that traditional close reading need not be at odds with machine reading, as each have “distinctive advantages and limitations,” and can “interact synergistically with one another.”²² Also, what data mining, computer-assisted analysis, and visualization offer literary interpretation can reveal unanticipated patterns within and among collections of texts that lead to new inquiries and new ways of reading texts more closely.²³ Many of the fears that critics voice in opposition to computer-based methods seem to stem from a refusal to acknowledge that the interpretive process is not unidirectional, but follows a sort of recursive loop. In Stephen Ramsay’s article titled “Stanley and Me,” he responds to a series of online essays that Stanley Fish published in the *New York Times* that not only critique the field of digital humanities, but name and criticize several prominent digital humanities scholars, Ramsay included.²⁴ In response, he states that the “idea that criticism only happens when you proceed from interpretative propositions to supportive patterns is, at best, a rhetorical dodge ... if one concedes that reading is a process whereby patterns urge interpretations and interpretations bring forth patterns (a formulation that implies no natural sequence or direction), then we are halfway to saying why you might want to generate a computational program from a critical question or generate an interpretation from a computation procedure. Why not do both? In fact, why not do both within tight, concentric loops for reading, hacking, thinking, and interpreting?”²⁵ It is my sincere belief, therefore, that computational analysis—whether analyzing one, two texts, or hundreds—works best when grounded in traditional humanist analysis. Machine reading may detect subtle patterns within a text or across a large corpus, but those findings should always be carefully scrutinized

20 Proponents of data mining and computer-based analysis, such as Franco Moretti, Matthew Jockers, Alan Liu, Johanna Drucker, Katherine N. Hayles, and Stephen Ramsay, among others, have drawn criticism from those who are skeptical about the digital humanities in general and computational analysis in particular—most notably, Stanley Fish who wrote a series of blog posts questioning the merits of this emerging field of study. See Moretti, *Distant Reading*; Moretti, *Graphs, Maps, and Trees*; Ramsay, *Reading Machines*; Ramsay and Rockwell, “Developing Things”; Drucker, “Humanistic Theory and Digital Scholarship”; Hayles, *How We Think*; Jockers, *Macroanalysis*.”

21 Kirsh, “Technology is Taking Over English Departments.”

22 Hayles, *How We Think*, 74.

23 Kirschenbaum, “Poetry, Patterns, and Provocation: The Nora Project.”

24 Fish, “The Old Order Changeth”; Fish, “The Digital Humanities and the Transcending of Mortality”; Fish, “Mind Your P’s and B’s.”

25 Ramsay, “Stanley and Me.”

and followed up by a close reading. In my *Beowulf-Blickling* project, for example, I used lexomic cluster analysis to look for patterns in my target texts and then switched to a formula density analysis that required a word-level close reading of the target passages in order to explain the presence of certain anomalies that the lexomic method detected. That project demonstrates how computational analysis supplemented, but never took precedence over my close reading of the text.

Middle English: Lexomics, Voyant, and Chaucer's *Canterbury Tales*

After experimenting with digital tools in Old English, I turned my attention to one of the most-analyzed Middle English texts, Chaucer's *Canterbury Tales*.²⁶ My goal was to use Lexomics and Voyant—another web-based text analysis tool that accommodates Old and Middle English text—in order to explore two questions concerning word usage in the *Canterbury Tales*. First, I performed a lexomic cluster analysis on the *General Prologue* to Chaucer's *Canterbury Tales* in order to see if Chaucer wrote differently about his female characters than he did about male characters; and secondly I used Voyant to look at which tales use the word *privetee* (private) and how often the word is paired with *apert* (public) in the text. In the second analysis, I wanted to find out whether *privetee* and *apert* are two linked ideas in a single phrase, or if the phrase takes the place of a single concept, forming a sort of idiomatic expression unique to Chaucer.

I began by generating a lexomic dendrogram for the *General Prologue* that included each of the pilgrims' portraits, and then produced word cloud visualizations that measured word frequency in portraits that stood out in any significant way. After separating each portrait into different plain-text documents, I ran a lexomic cluster analysis and produced the dendrogram in Figure 2.2.

I produced a word cloud visualization using Voyant for the Wife of Bath because the dendrogram revealed a close clustering of her portrait and the Prioress's. The visualization revealed that two feminine pronouns, *she* and *hir*, appeared frequently in her portrait, as pictured below in Figure 2.3.

I generated a similar word cloud and word frequency list for the Prioress's portrait and found that it was also dominated by feminine pronouns, as pictured in Figure 2.4.

The strong lexical connection between the portraits of the Prioress and the Wife of Bath resides primarily in their shared feminine pronouns. The two Voyant word collocation visualizations²⁷ below confirm this information. Fig. 2.5 below visualizes both portraits and includes pronouns that are shared by both characters. The feminine pronouns *she* and *hir* appear frequently in both portraits and are the reason that the Wife of Bath and Prioress text-chunks clustered together in the above dendrogram.

²⁶ For this project, I used an online machine-readable version of F. N. Robinson, ed. *The Works of Geoffrey Chaucer*, 2nd. ed. I used this version of the *Canterbury Tales* because Robinson used the original spelling in the manuscripts and clearly documented any emendations he made to the text in the Textual Notes section.

²⁷ Voyant collocation viz may be accessed at: <http://docs.voyant-tools.org/tools/links/>. The Voyant collocation visualization tool represents the frequency of a word by the relative size of

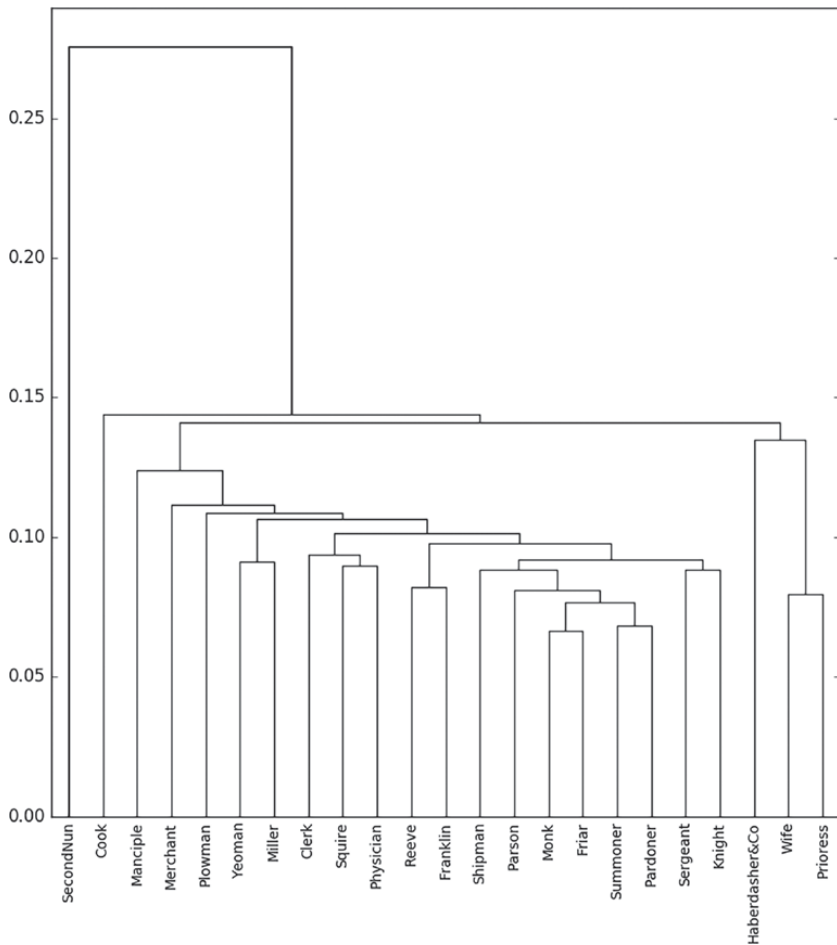


Figure 2.2. Dendrogram—*Canterbury Tales: General Prologue* portraits.

Figure 2.6 below visualizes the two portraits with all function words removed. Once the function was removed, including the shared feminine pronouns, *Ful* seems to be the only word the Wife of Bath and Prioress portraits share in common.

After generating a second dendrogram for all of the portraits in the *General Prologue* that excludes all function words, the Wife of Bath and Prioress portraits clustered at a significant distance from one another, reaffirming that the connection between these

the term. Mousing over any word in the online version of the collocation visualization will display additional information about the word's occurrence within a text or across a corpus, including frequency and the number of documents the word appears in.

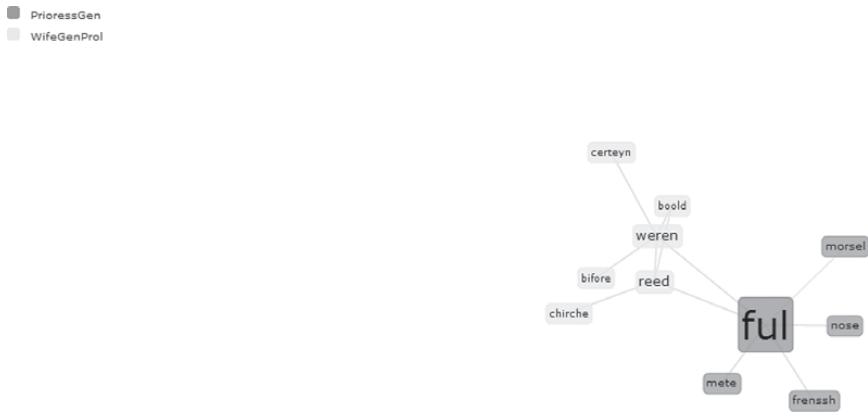


Figure 2.6. Combined Wife of Bath and Prioress Portraits excluding function words (Voyant Word Collocation tool).

two characters lies primarily in their shared feminine pronouns. While this analysis is still in its preliminary stages, and is in no way revolutionary, initial findings suggest that similarities and differences in the ways that Chaucer presents his female pilgrims in the *General Prologue* may exist at the surface level—in his use of function words—rather than in distinctions he makes based on gender differences. If anything, this exercise serves as a caution for those who would use a digital clustering analysis of these profiles to make claims, for instance, about Chaucer’s understanding of gender roles. Because these profiles do not cluster together when gendered pronouns are removed, the most we can say about the passages based on a cluster analysis is that the subjects of both of these profiles are likely female. While a closer reading might bear more fruit in an exploration of gender roles, in this case, any conclusions about gender roles based on a clustering analysis would be methodologically unsound. Using Lexomics and Voyant in this way is an excellent way to supplement research and demonstrates the different uses of out-of-the box tools that are user-friendly and accommodate both Old and Middle English texts. Also, while much of the word-counting can be done by hand, these tools accomplish the task with greater speed and accuracy and visualize that information in new and interesting ways.

The second part of my analysis used Voyant to examine the word *privetee* and its cognates in Chaucer’s *Canterbury Tales*, and took as its starting point Edward Costigan’s article titled “‘Privetee’ in the *Canterbury Tales*.”²⁹ Costigan analyzes how and where Chaucer uses the word *privetee* and its cognates: *privee*, *pryvee*, *privitee*, *pryvetee*, *pryvely*, and *prively*, and argues that the word can mean “private,” “intimate,” “secret,” and “with-drawn.” In the broadest sense, the word functions as a division between the private thoughts and affairs of individuals and their public actions within their community.

²⁹ Costigan, “‘Privetee’ in the *Canterbury Tales*,” 217–30.

Privetee and its cognates appear in several tales, sometimes collocating with the word *apert*. The phrase *privee and apert* is used “to support the idea of a perfect unity of personal honour and public reputation, the two becoming one ... inner motive and outward behaviour should be in harmony.”³⁰ This idea, however, is complicated by the question of where and how the distinction between private and public life was made in the Middle Ages. Costigan points out that in Chaucer’s time, people lived their lives more publicly and that Chaucer’s use of the word “private” sometimes “shades into ‘secret’ and ‘treacherous.’”³¹ The word “private” was used to describe the Reeve’s secret stash of money as well as Arcite’s secret plans to disguise himself in order to gain access to Emily. The wicked sergeant in the “Clerk’s Tale” was described as a “privee man” and the very deceitful Nicholas in the “Miller’s Tale” was “sleigh and ful of privee.”³² According to Costigan, the word appears most frequently in the “Wife of Bath’s Tale,” the “Shipman’s Tale,” and the “Miller’s Tale,” where privacy and concealment signal not only difficulties in marital relationships, but problems in the community as a whole. The Wife emphasizes the knight’s concealment of his marriage to the loathly lady rather than celebrating it publicly. Her tale, which focuses on marriage—a social event that unites public and private spheres—uses the detail of a private wedding to highlight not only how wrong the union was, but the unfairness of a community that would force a man into a miserable marriage because of an oath.³³ The “Shipman’s Tale” and the “Miller’s Tale” also use the word *privetee* and its various cognates to highlight the difference between public acts and private acts that destroy relationships and compromise the integrity of the social fabric of a community. The ideal, therefore, seems to be a balance between private honour and public action—a sort of concord between *privee* and *apert*.

In order to gain a better understanding about where *privetee* and its cognates appear in *Canterbury Tales* and how frequently *privee* and its variant, *pryvee* collocate with *apert*, I used Voyant to calculate word frequency and collocation patterns. The analysis revealed that, contrary to Costigan’s findings, it is in the “Parson’s Tale, rather than in the “Wife of Bath’s Tale,” the “Shipman’s Tale,” or the “Miller’s Tale” that *privee* and its variants appear with the greatest frequency. Furthermore, *pryvee* and *pryvee* only collocate with *apert* three times in the entire corpus—once in the “Squire’s Tale” and twice in the “Wife of Bath’s Tale.”

Using Voyant’s Reader and Context tools to pinpoint the exact location of each instance of *privetee* and its variants in the *Canterbury Tales*, I then performed a close reading and found that the term “private” and its variants appear in each tale at climactic moments when there is some sort of moral conflict or tension between private, dubious actions and honest activities performed in public spaces. I found that this tension exists in the either the interior spaces (or the conscience) of a single character, or in the social

30 Costigan, 217–18.

31 Costigan, 220.

32 Costigan, 220, 225.

33 Costigan, 223.

spaces shared by various characters in a community. Also, despite the infrequent pairing of *privee* and *apert*, tension between public and private is always implied wherever *privitee* and its variants appear in the text.

A closer look at the use of the word *prively* in the “Nun’s Priest’s Tale” illustrates that Chaucer’s use of the word and its related forms is consistent in all of his works. The “Nun’s Priest’s Tale” references the word *prively* when Chauntecleer attempts to convince Pertelote that dreams can convey important information. He tells her a story about the shade of a murdered man who visits his friend in a dream and tells him that he was killed by an innkeeper for his money, and also reveals where to find his body: “A carte ful of dong ther shaltow se / In which my body is hid ful prively” (3018–19). Once the body is found, the innkeeper’s crime is exposed. What was once private and hidden is seen in the clear light of day and dealt with by the community:

O blisful God, that art so just and trewe,
Lo, how that thou biwreyst mordre away!
Mordre wol out; that se we day by day.
Mordre is so wlatson and abhomynable
To God, that is so just and resonable,
That he ne wol nat suffre it heled be,
Though it abyde a yeer, or two, or thre.
Mordre wol out, this my conclusioun.”

[O blissful God, who art so just and true,
Lo, how that thou always reveal murder!
Murder will out; we see that every day.
Murder is so loathsome and abominable
To God, who is so just and reasonable,
That He will not suffer it to be hidden,
Though it may remain a year, or two, or three.
Murder will out, this is my conclusion.]

(305 0–57)

Interestingly, *prively* is not used again in the tale. One would expect it to appear in the description of the crafty fox who tricks and nearly eats Chauntecleer, as the fox, like Nicholas in the “Miller’s Tale,” is a trickster. However, unlike Nicholas, he is not associated with the words *privitee* or *privee*. *Privitee* and its cognates appear in the “Miller’s Tale” in connection with Nicholas and Alison’s secret sexual transgressions. It appears that *prively* functions in the “Nun’s Priest’s Tale” in the same way that it does in the “Miller’s Tale.” Secret deeds that harm and destroy are eventually exposed to public view and corrected by the community. The fox is crafty, but his seduction of Chauntecleer through false flattery is not a completely hidden act. In Chauntecleer’s tale about the man murdered by the innkeeper, balance is restored between *privitee* and *apert* once the crime is brought to light. Private thoughts and secrets are exposed and punished so that only noble thoughts and acts remain. This is a running theme in most of the tales that is underscored by the presence and distribution of *privitee* and its cognates.

My brief analysis of *privitee* and its variants in Chaucer's *Canterbury Tales* illustrates how a tool like Voyant can provide an accurate and time-saving roadmap, identifying areas worthy of further investigation in Old and Middle English texts. Although some still fear that digital analysis involves reducing great literary works to quantized units of machine-digestible data to be processed by tools that spit out answers to eager researchers who only have to click a submit button, anyone who has used digital tools in humanities research knows that digital methodologies work best when they supplement traditional research methods. The Lexomics FAQ page most eloquently sums this idea when responding to the humorous inquiry: "Will lexomics replace me as a scholar?" The response is as follows: "Nope. Lexomics, like most computational text mining tools, are what John Burrows calls a 'middle game' technique. You (the scholar) have much scholarship to do before running computational tools (for example, collecting texts, forming hypotheses as to where you might segment your texts, etc.) and then following the use of the tools, you have much work to do after to interpret the results and form new hypotheses. The computer is just a tool 'in the middle,' albeit a very powerful tool that today's scholars of texts want in their arsenal."³⁴ In other words, computer-aided analysis cannot replace a close reading of the text itself, and rigorous scholarship is always needed to interpret the results of such analyses.

In this chapter, I discussed only two of the many digital tools that are now available to humanist scholars. As new digital research tools are always being developed and existing ones are continually updated and improved, I check DiRT and TAPoR regularly. I have since discovered and used other tools that are compatible with Old and Middle English texts, such as Gephi for literary network analysis,³⁵ KH Coder for quantitative content analysis and text mining,³⁶ and Visualeyes for spatio-temporal mapping of events in texts,³⁷ to name a few. The value of digital tools lies in their ability to point to areas worthy of further investigation, automate certain processes with greater efficiency—such as word counting and tracking—and to visualize that information in new and thought-provoking ways. Incorporating digital tools into my scholarship has changed how I do research and write about my findings. I have accepted that projects that involve computer-assisted analysis are sometimes open-ended and have to occasionally account for failure, as they depend on tools that are continually changing and being updated. As such, I have learned that documenting every step of the research process is crucial in order to establish clarity and maintain transparency for my readers and collaborators. With more peer-reviewed scholarship featuring digital tools in medieval studies, scholars will have more to draw upon in terms of established good practices. And as more digital scholarship becomes available in the field, perhaps the critics of computer-assisted scholarship will come to see that digital methods are not simply quantize, click, and conclude.

³⁴ Drout et al., "Lexomics FAQ."

³⁵ Bastian, Heymann, and Jacomy, "Gephi."

³⁶ Higuchi, KH Coder.

³⁷ Ferster, VisualEyes.

Bibliography

- Bastian, M., S. Heymann, and M. Jacomy. "Gephi: An Open Source Software for Exploring and Manipulating Networks." Gephi, 2009. <https://gephi.org/>.
- Brown, Carleton. "Beowulf and the *Blickling Homilies* and Some Textual Notes." *PMLA* 53 (1938): 905–16.
- Costigan, Edward. "'Privetee' in the *Canterbury Tales*." *Studies in English Literature* 60 (1983): 217–30.
- Drout, Michael. "Lexomics FAQ." Lexomics. <http://wheatoncollege.edu/lexomics/faq/#cite>.
———. *Tradition and Influence in Anglo-Saxon Literature: An Evolutionary, Cognitivist Approach*. New York: Palgrave Macmillan, 2013.
- Drout, Michael, Michael J. Kahn, Mark D. LeBlanc, and Christina Nelson. "Of Dendro-grammatology: Lexomics Methods for Analyzing Relationships among Old English Poems." *Journal of English and Germanic Philology* 110 (2011): 301–36.
- Drout, Michael, Yvette Kisor, Leah Smith, Allison Dennett, and Natasha Piirainen. *Beowulf Unlocked: New Evidence from Lexomic Analysis*. New York: Palgrave Macmillan, 2016.
- Drucker, Johanna. "Humanistic Theory and Digital Scholarship." In *Debates in Digital Humanities*, edited by Matthew K. Gold, 85–95. Minneapolis: University of Minnesota Press, 2012.
- Ferster, Bill. VisualEyes. Sciences, Humanities & Arts Network of Technological Initiatives. University of Virginia, 2016. www.viseyes.org/viseyes.htm.
- Fish, Stanley. "Mind Your P's and B's: The Digital Humanities and Interpretation." *New York Times*, January 23, 2012.
- . "The Digital Humanities and the Transcending of Mortality." *New York Times*, January 9, 2012.
- . "The Old Order Changeth." *New York Times*, December 26, 2011. http://opinionator.blogs.nytimes.com/2011/12/26/the-old-order-changeth/?_r=0.
- Hayles, Katharine N. *How We Think: Digital Media and Contemporary Technogenesis*. Chicago: University of Chicago, 2012.
- Higuchi, Koichi. K H Coder, 2001. <http://khc.sourceforge.net/en/>.
- Jockers, Matthew L. *Macroanalysis: Digital Methods and Literary History*. Champaign: University of Illinois Press, 2013.
- Kirschenbaum, Matthew G. "Poetry, Patterns, and Provocation: The Nora Project." *The Valve*, 2006. www.thevalve.org/go/valve/article/poetry_patterns_and_provocation_the_nora_project/.
- Kirsh, Adam. "Technology is Taking Over English Departments: The False Promise of the Digital Humanities." *New Republic*, May 2, 2014. <https://dhknowledgegroup.wordpress.com/2014/05/13/technology-is-taking-over-english-departments-the-false-promise-of-the-digital-humanities/>.
- Klaeber, Frederick, ed. *Klaeber's Beowulf and the Fight at Finnsburg*. 4th ed. Edited by R. D. Fulk, Robert E. Bjork, and John D. Niles. Toronto: University of Toronto Press, 2008.
- Kleinman, Scott, Mark D. LeBlanc, and Michael Drout. "Lexomics Tools," 2015. <http://wheatoncollege.edu/lexomics>.

- Malone, Kemp. "Grendel and His Abode." In *Studia Philologica et Litteraria in Honorem*, edited by L. Spitzer and K. L. Selig, 297–308. Bern: Francke, 1958.
- Moretti, Franco. *Distant Reading*. New York: Verso, 2013.
- . *Graphs, Maps, and Trees: Abstract Models for a Literary History*. London: Verso, 2005.
- Morris, Richard, ed. *The Blickling Homilies of the Tenth Century*. London: Early English Text Society, 1880.
- Ramsay, Stephen. *Reading Machines: Towards an Algorithmic Criticism*. Champaign: University of Illinois Press, 2011.
- . "Stanley and Me" (blog), November 8, 2012. <http://stephenramsay.us/text/2012/11/08/stanley-and-me/>.
- Ramsay, Stephen, and Geoffrey Rockwell. "Developing Things: Notes Toward an Epistemology of Building in the Digital Humanities." In *Debates in Digital Humanities*, edited by Matthew K. Gold, 75–84. Minneapolis: University of Minnesota Press, 2012.
- Robinson, F. N., ed. *The Works of Geoffrey Chaucer*. 2nd ed. Houghton Mifflin: Boston, 1957. www.unc.edu/depts/chaucer/chtexts.htm.
- Rockwell, Geoffrey, Kristen C. Uszkalo, and Stéfan Sinclair. The Text Analysis Portal for Research (TAPoR), 2015. http://tapor.ca/pages/about_tapor.
- Scragg, D. G. "The Corpus of Vernacular Homilies and Prose Saints." *Anglo-Saxon England* 8 (1979): 223–77.
- Sinclair, Stéfan, Geoffrey Rockwell, and the Voyant Tools Team. Voyant Tools (web application), 2012. <http://docs.voyant-tools.org/about/>.
- Spiro, Lisa. Digital Research Tools Directory (DiRT), 2015. <http://dirtdirectory.org/contribute>.
- Torabi, Katayoun. "Two New Approaches to Exploring Monstrous Landscapes in *Beowulf* and *Blickling Homily XVII*." *Essays in Medieval Studies* 31 (2015): 165–82.

Chapter 3

PROJECT PARADISE: A GEO-TEMPORAL EXHIBIT OF THE HEREFORD MAP AND THE BOOK OF JOHN MANDEVILLE

ALEXANDRA BOLINTINEANU

Introduction

“OF PARADISE I CANNOT speak properly, for I have not been there; and that I regret,” writes the pilgrim-narrator in *The Book of John Mandeville*.¹ In medieval Christian thought, Paradise is both the lost Earthly Paradise of Eden and the Heavenly Jerusalem yet to come—a place on earth, in the world’s uttermost East, and a region of the afterlife; the first, lost home of humankind, and the last home of righteous souls with God.² Project Paradise is a digital project that brings into conversation two late medieval representations of the Earthly Paradise: the thirteenth-century Hereford mappa mundi (world map) and the fourteenth-century *Book of John Mandeville*.³ Both the mappa mundi and Mandeville place the Earthly Paradise in the uttermost East of the world.⁴ This project takes the Mandeville’s references to Paradise and attempts to place them on the map. Initially, I thought the exhibit would trace the continuities between the mappa mundi and Mandeville, the ways Hereford and Mandeville both draw on biblical narrative and the encyclopaedic tradition to create a geography of the East.⁵ But what I could *not* map was more striking than what I could: many of the allusions to Paradise in Mandeville have no place on the Hereford map.

This digital exhibit reveals that geography in Mandeville is far more pervasively oriented towards Paradise than the Paradise-topped mappa mundi. The Hereford map draws a clear visual boundary around the Earthly Paradise. *The Book of John Mandeville*, on the other hand, makes the boundary porous. Its inaccessible Paradise infiltrates the entire geography of the East. In Hereford, Paradise is remote and enclosed; in Mandeville,

1 *The Book of John Mandeville*, eds. Kohanski and Benson.

2 Akbari, *Idols in the East*, 248–49.

3 This study references the Hereford Mappa Mundi in Westrem, *The Hereford Map*, and *The Book of John Mandeville*, eds. Kohanski and Benson, a single-manuscript edition of British Library MS Royal 17 C. xxxviii. Underlying the digital exhibit is a photograph of the Hereford Mappa Mundi, provided by Hereford Cathedral, The Cathedral Library and Archives, and used with their generous permission.

4 For overviews of Paradise’s place in medieval thought, see Akbari, 20–66, 248–49; Scafi, *Mapping Paradise*, 32–61, 84–159.

5 Edson, *Mapping Time and Space*, 18–144; Zumthor, *La Mesure du Monde*; and Woodward, “Medieval Mappaemundi,” 286–369. See also Albu, *The Medieval Peutinger Map*, esp., 31–41, for the contrast between late Roman and early Christian conceptualizations of world geography.

Paradise is a constant yet emphatically incomplete presence throughout the imagined East, surfacing through physical relics and remains, customs and legends, and fragmentary reflections.

Projecting a digital archive of Mandeville's Paradise references onto the Hereford map thus accomplishes three goals. It displays the relationship between text and map in a single image, as printed editions of map and text would not allow.⁶ It attempts to reconstruct medieval spatial poetics, translating and mediating between map and narrative, between the intersecting categories of space, time, and story.⁷ And it invites us to reflect on the role of digitality in humanities scholarship, arguing that the digital platforms through which we access the medieval archive act as modes of knowledge production, as they shape forms of reading and understanding historical materials.

The methodological basis of this study is the notion of deformance, a portmanteau term coined by Lisa Samuels and Jerome McGann, a combination of "deform" and "performance."⁸ Deformance is a strategy of interpreting a literary text by invading it and changing its shape: by reading a poem backwards line by line, or by turning it to prose; by presenting poems as prose and prose as poems, or by reading exclusively a literary text's nouns or its verbs. Deformance makes a familiar text strange, reopening its interpretive possibilities. Deformance of textual objects need not be limited to reading or punctuation. As Johanna Drucker and Bethany Nowviskie note, deformance can take the shape of speculative visualization, interpretive thinking about text and culture through the creation of digital artefacts.⁹ Thus, Project Paradise practices deformance by pulling Mandeville's references to Paradise out of their immediate narrative context and projecting them on the map. But in so doing, the project restores a wider historical context, using the Mandeville passages to re-enact medieval spatial storytelling practices. For Neatline (the Omeka plugin used for Project Paradise) and medieval world maps are

6 The standard Hereford edition is Westrem, *The Hereford Map*. It presents the Hereford Map as follows: first, as a series of one-page panels that reproduce sections of the map at full size, with map legends numbered and keyed to transcriptions appearing later in the book; second, as a fold-out insert that reproduces the full map, at a scale of approximately 1:2.5 (the reproduction map is approximately 20 inches/50 cm tall; the Hereford map itself is approximately 52 inches/132 cm tall). As reviewers have noted, Westrem's seminal study is difficult to read, as the reader must navigate between numbered image, numbered transcriptions in a different section of the volume, and whole image in the appendix (Tomasch, "Reviews: *The Hereford Map*, Scott D. Westrem," 922). Greenlee, in his *Mapping Mandeville Project*, annotates a 1989 facsimile copy of the Hereford Map with places named in *The Book of John Mandeville*: that is, he identifies and annotates places that the two texts have in common. My project, by contrast, is smaller in scope: it compares the two documents along a common theme.

7 See Woodward, "Reality, Symbolism, Time, and Space." For a comprehensive survey of medieval world maps, see Woodward, "Medieval Mappaemundi," 286–370. For a discussion of the narrative and theological functions of medieval maps—their spatial poetics, as it were—see Woodward, "Reality, Symbolism, Time, and Space," 510–21.

8 Samuels and McGann, "Deformance and Interpretation," 25–56.

9 Drucker and Nowviskie, "Speculative Computing."

interestingly akin to each other: both are platforms that transform space into time, time into space.

Digital Mapping

In literary and historical studies, digital mapping has developed significantly over the last decade.¹⁰ As digital mapping platforms become increasingly accessible, both in cost and user-friendliness, “[v]isualization has become part and parcel of the research process. We think through our research through seeing data on maps.”¹¹ GIS (geographical information system) software allows scholars to visualize their data on maps, by assigning geographical coordinates (longitude and latitude) to historical and literary data, then uploading the data into software such as Palladio, Carto, or even ArcGIS, and finally visualizing their data (literary texts or historical events) as points on a digital map.

Reviewing digital work on historical maps, Martin Foys distinguishes between three approaches.¹² The first category is that of “modern, measured, plotted and/or georectified spaces”—historical maps that are or have been made as accurate as possible and annotated with historical data about that space. For example, Project DECIMA studies sixteenth-century Florence. DECIMA projects data from the 1563 Florentine census onto an “exceptionally accurate and detailed” map of Florence from 1584, only two decades after the census. DECIMA reveals house-by-house data about Florence’s inhabitants: where Florentines lived, where they worked, where they worshipped. In so doing, the project examines sixteenth-century Florentines’ “social networks, economic currents, and the sensory life of Florence.”¹³ GIS projects track not only populations, but texts and their transmission. For example, David J. Wrisley’s project, French of Italy, map, centres of literary production of French documents on the Italian peninsula in space and time. The map is linked to a timeline and dynamically updates the centres of literary production “active” at a given time, as users progress along the timeline. While each of these projects uses different digital platforms, they share the aim of visualizing historical and literary information in the spaces where it occurred. Similar in its use of modern georectified maps, but different in its focus on literary rather than historical geographies, David Joseph Wrisley’s project Visualizing Medieval Places (VMP) tracks “textual references made to different kinds of place names at different scales within sustained prose or poetic narrative—landmarks, settlements, regions, and countries—real and

10 For an overview of recent works, see Wrisley, “Spatial Humanities.” Among other works on spatial methodologies, Wrisley highlights Hillier and Knowles, *Placing History*; Bodenhamer, Corrigan, and Harris, *The Spatial Humanities*; Gregory and Geddes, *Toward Spatial Humanities*; Montello, Grossner, and Janelle, *Space In Mind*. See also Bonnell and Fortin, eds., *Historical GIS Research in Canada*.

11 Bonnell and Fortin, *Historical GIS Research*, 97.

12 Foys and Bradshaw, “Developing Digital Mappaemundi.”

13 DECIMA website.

imaginary.”¹⁴ As Wrisley notes, VMP collects this geographic information as it occurs in medieval French texts composed between the eleventh and the fifteenth centuries, and then visualizes this information through modern, georectified maps. Rather than track the places where the texts themselves are produced, VMP tracks the places the texts themselves evoke, mapping these places by literary tradition, genre, and time period, and tracing the contours of the world’s geography as imagined across medieval French texts.

The second category of digital mapping projects, according to Foys, is that of historical maps annotated with external content, such as images, sound files, literary texts. For example, the Map of Early Modern London project links “encyclopedia-style articles, scholarly work, student work, editions, and literary texts” to related locations on the 1578 Agas map of London.¹⁵

The third category—launched, Foys argues, by his own Digital Mappa Mundi project—is that of historical maps annotated with their own content; that is, searchable digital transcriptions and editions of historical maps. Project Paradise is located in the overlap between Foys’s second and third categories. It annotates a map with content that is external to the map, but that engages with the map’s ideology, its intellectual goals of functioning as a cosmography of the world, its project of building the world’s history and theology on spatial foundations. Project Paradise participates in the map’s spatial poetics, its mediation between story and space, between geography and history.

Why Compare Hereford and Mandeville?

In the evocative words of John B. Friedman, medieval world maps and *The Book of John Mandeville* are “gigantic detached illustrated leaves of the book of the world”:¹⁶ accounts of the world that depict, onto a geographical framework, the world’s history, its population, and its theological purpose. In the world they picture, Jerusalem is at the centre; the Earthly Paradise endures in the uttermost East; events from classical and biblical history are layered onto geographical landmarks; and monstrous or marvellous creatures inhabit the world in their diversity.¹⁷ And in this project of describing space, its inhabitants, its layers of history, and its network of theological meanings, both the world maps and Mandeville draw on similar traditions: on the classical *cosmographia* tradition, exemplified by Pliny and Solinus, and later by Orosius and Isidore of Seville; on the biblical tradition; on the *mirabilia* tradition; and on the encyclopaedic tradition of the thirteenth and (for Mandeville) fourteenth century.¹⁸

¹⁴ Wrisley, “Locating Medieval French,” 145–69.

¹⁵ Jenstad, ed., “The Map of Early Modern London.”

¹⁶ Friedman, “Cultural Conflicts in Medieval Maps,” 64.

¹⁷ For a discussion of Jerusalem’s centrality in mappaemundi, see Woodward, “Medieval Mappaemundi,” 301–2; Edson, *Mapping Time and Space*, 59–65. For a discussion of Jerusalem’s centrality in Mandeville, see Akbari, *Idols in the East*, 51–59, and Higgins, “Defining the Earth’s Center,” 29–53.

¹⁸ On Mandeville’s sources, see Deluz, *Le livre de Jehan de Mandeville*, 428–91. For the sources of the mappaemundi, see Westrem, “Making a Mappamundi,” 19–33 (esp. 24); Westrem, *The Hereford Map*, xxviii–xxxvii and notes 43–59. See also Edson, *Mapping Time and Space*, 59–61.

Space is a canvas on which the mapmakers and the Mandeville narrator draw a portrait of the world—with comparable intellectual goals, and from a comparable treasury of interwoven traditions. In the so-called “papal interpolation,” even the Mandeville-narrator connects his narrative to a *mappa mundi*:

And for as moche as many man troweth noght but that they se with her owen eye other that they may conseve with her kyndely witte, therfore Y maked my way in my tornyng homwarde to Rome to showe my book to the holy fader, the pope, and telle to hym mervayles whoch Y hadde y-seye in diverse contrees, so that he, with his wise consayl wolde examine hit with diverse peple that beth in Rome. For ther beth in Rome evermore men dwellynge of diverse nacions of the worlde. And a litel tyme after, when he and his conseil hadde examined hit al thorgh, he sayde me for certayn that all was soath that was therynne. For he sayde he hadde a book ppon Latyn that conteyned that and moch more, after which book the *mappa mundi* ys y-maked, which book he shewed to me. And therfore the holi fader, the pope, hath ratified and confermed my book in alle poyntes.

[And because many people believe nothing except what they see with their own eyes or what they can conceive with their natural wit, therefore I made my way in my turning homeward to Rome to show my book to the holy father, the pope, and tell him marvels which I had seen in different countries, so that he with his wise council would examine it with different people who are in Rome. For there are always in Rome people of different nations of the world. And a little time after, when he and his council had examined it all through, he told me for certain that all was true that was in it. For he said he had a book in Latin that contained that and much more, after which book the *mappa mundi* is made, which book he showed to me. And therefore the holy father, the pope, has ratified and confirmed by book in all points.]

(ll. 2838–48)

The Mandeville-narrator invokes both papal and geographical authority, linking his own narrative to a “book ppon Latyn ... after which book the *mappa mundi* ys y-maked,” a book that he claims confirms “alle poyntes” of his own narrative. This episode does not appear in the French texts of the *Book*. Rather, it is thought that it was added into the *Book* sometime after its translation into English, possibly not long before 1400.¹⁹ However, the papal interpolation appears in all complete English texts of the *Book*—except for the two English manuscripts of the Bodley Version (Bodleian Library MSS Rawlinson D 99 and eMusaeo 116)—as well as in a single Latin text (Durham University MS Cosin V. iii. 7), linking the Mandeville-narrative quite explicitly to the *mappa mundi* tradition. Moreover, the “textual isotope” of the *Book* on which the current project is based—namely, British Library MS Royal 17 C. xxxvii (hereafter MS Royal 17 C. xxxvii)—contains not only the papal interpolation, but also a tiny picture of “the compasse of the

¹⁹ Kohanski and Benson, Notes (ll. 2838–48), *The Book of John Mandeville*; Seymour, “The Origin of the Egerton Version,” 159–69; Higgins, *Writing East*, 254–60.

erthe" (the bounds of the earth), a damaged, upside-down T-O map.²⁰ Within the text and in its margins, this particular text of *The Book of John Mandeville* is deeply involved with the geographical traditions of the mappae mundi.

Given these commonalities, it is "almost a cliché to compare the world depicted in *The Book of John Mandeville* to the world depicted pictorially on the mappa mundi."²¹ So far, however, the larger critical conversation has compared two traditions, Mandeville and medieval world maps, in general and synthetic terms. That is, it has compared mappae mundi—diverse in shape, granularity, detail, narrative structure—to the Mandeville multi-text, whose versions are also highly diverse in their ideological aims, narrative structure and content, form, and representation of space.²² By contrast to this general and synthetic approach, I want to focus the comparison on two specific instances: the Hereford map, the largest, most detailed medieval world map that still survives; and *The Book of John Mandeville*'s so-called Defective Version, the version of the text most widespread in England, as it appears in MS Royal 17 C. xxxvii.²³ This comparison enables me to collect specific textual data and project it against the map, transforming the more general comparison of two genres into a geo-temporal exhibit of a textual archive, in order to study Hereford's and Mandeville's spatial poetics. The comparison of a specific text and a specific map narrow my argument's field of vision; but at the same time, the comparison allows me to test more synthetic insights into Mandeville's spatial poetics against close-grained textual and visual evidence.

Project Paradise²⁴ therefore contributes to two scholarly conversations. First, it engages in the conversation around digital mapping. Second, it contributes to

20 British Library Royal 17 C. xxxviii f. 41v, with the Middle English caption limitingly translated in the British Library Illuminated Manuscript Catalogue entry as "compass of the world."

21 Akbari, *Idols in the East*, 66. For comparisons between Mandeville and the mappae mundi, see among others Higgins, "Defining the Earth's Center," esp. 49–51, "The Book and the Maps"; Yeager, *Jerusalem*. But see also Moseley, "Behaim's Globe," 89–91 for a comparison between the narrative of Mandeville and Martin Behaim's Globe, the earliest extant globe, made at Nuremberg in 1492, and in part based on Mandeville's description of the world. Combining map and globe as models for conceptualizing Mandeville's geography, Christiane Deluz argues that the world geography evoked by the *Book* is—in the *Book*'s proclamation of Christianity's all-embracing reach—"la transposition écrite de ce que signifiait la mappemonde d'Ebstorf en enserrant la terre dans les bras du Christ" (the written transposition of what the Ebstorf world map signifies in enclosing the Earth in the embrace of Christ), 19–21; and also productively compared to Martin Behaim's globe, 88–89; Edson, *The World Map, 1300–1492*, 68. See also Greenlee, *The Mapping Mandeville Project*, which "seeks to combine Mandeville's text with one of the largest and most famous of the mappamundi—the Hereford Map—in an attempt to use the two media together to make them more individually comprehensible" (Greenlee, "About").

22 Akbari, *Idols in the East*, 66.

23 The focus on a single manuscript of *The Book of John Mandeville*—among a tradition whose "textual isotopes" (that is, versions and manuscript variants), in Iain MacLeod Higgins' wording, vary so widely—is persuasively advocated by Higgins, *Writing East*, 17–28, and by Kohanski, ed., Introduction, *The Book of John Mandeville*, x–xiv.

24 Due to restrictions on the Hereford Map images, Project Paradise is currently behind a password wall. Please email alexandra.bolintineanu@utoronto.ca for access information.

the conversation around Mandeville's representation of the world and its theological centres of gravity—the swing of its narrative pendulum between Jerusalem and the Earthly Paradise.²⁵ The project traces a spatial poetics of wonder that represents Paradise as an elusive, fragmentary presence, scattered through the geography of the world. Thus the poetics of wonder that surrounds Paradise in *The Book of John Mandeville* is a poetics of unknowing, of spatial indeterminacy. This aligns Mandeville more strongly not with the cartographic tradition but with textual traditions—specifically, with romances, visions of the afterlife, and marvel narratives. In Old and Middle English, these traditions consistently portray spaces of wonder—places like the underground pygmy kingdom under Wales, in Gerald of Wales *Itinerarium Kambriae*; the Faerie realm in *Sir Orfeo*; or the Green Chapel in *Sir Gawain and the Green Knight*—as places that are secret, hidden, map-resistant, unknown to or unreachable by human-kind.²⁶ In this chapter, I will first show the digital exhibit that maps (and fails to map) Mandeville's Paradise references onto the Hereford mappa mundi. Then I will demonstrate how the conversation between Hereford and Mandeville clarifies the thematics of Mandeville's references to Paradise. Finally, I will talk about the broader theoretical implications of the project. In its choice of platform and configuration of data, Project Paradise moves digital mapping away from the georeferenced and quantified, into a more narrative kind of space: the kind of space created both by mappae mundi and by geo-temporal exhibits, a space made not from georeferenced coordinates but rather from intertwining stories.

The Exhibit

To construct the digital exhibit, I first built an archive of textual references to Paradise from *The Book of John Mandeville*. Then I used Mandeville's own geographical clues as well as Scott Westrem's edition and transcription of Hereford to place the Mandeville references onto a digital surrogate of the Hereford map.

Project Paradise's archive of textual references to Paradise from *The Book of John Mandeville* resides in Omeka, a content management system for digital archives and collections developed by the Roy Rosenzweig Center for History and New Media at George Mason University. Using Omeka, scholars can curate collections of digital items (manuscript facsimiles, photographs of artwork, texts, videos, sound recordings, events, performances); describe these items with rigorous metadata (Dublin Core standard) for each item, ensuring the collection's discoverability and interoperability with aggregators and library catalogues; and build multimedia-rich narratives from

²⁵ For a discussion of Jerusalem's centrality in Mandeville, see Akbari, *Idols in the East*, 51–59, and Higgins, "Defining the Earth's Center," 29–53.

²⁶ For spaces of wonder in medieval romance, see Saunders, *The Forest of Medieval Romance*; for spaces of wonder in Old and Middle English narrative in general, see Bolintineanu, "Towards a Poetics of Marvellous Spaces."

these items, using and reusing them in desired contexts.²⁷ An Omeka archive is made up of Items, each described with the Dublin Core metadata schema. An internationally accepted metadata standard, Dublin Core is neither very specialized nor very granular: each curator must interpret Dublin Core for her own collection, deciding how to allocate such fields as “Creator,” “Contributor,” or “Description” of a resource.²⁸ In this project, each archive Item represents a “Mandeville” reference to Paradise from my chosen Mandeville manuscript, Mandeville’s references to Paradise, as they are rendered in one specific version of the *Book*, as that text appears in British Library MS Royal 17 C. xxxvii.²⁹

The current collection maps the Defective Version against the earlier French text, *Livre de merveilles*, the earliest and fullest version of the Mandeville multi-text.³⁰ In future iterations of the project, I plan to separate the strands, mapping each Paradise reference from each Mandeville variant as its own record, so as to enable me to track and map Paradise references from different manuscripts in comparison to one another. At present, however, each Mandeville reference tracks three strands: first, the Middle English text from MS Royal 17 C. xxxvii; second, the Old French text from the *Livre de merveilles*, in Christiane Deluz’s critical edition;³¹ and third, a translation of the Middle English into Modern English. This organization allows readers to focus on MS Royal 17 C. xxxvii but at the same time compare its Paradise references to the Continental tradition.

Drawing on this Omeka archive of Items from the Mandeville text, I built a digital map using the Omeka plugin Neatline. Neatline is an open-source software platform, created by the Scholars’ Lab at the University of Virginia Library. Released as a suite of modular tools for the Omeka content management system, Neatline allows scholars and curators to create visual narratives by combining maps, timelines, and collections of artefact.³² In less technical terms, Neatline lets its users put time and story on a map; in my case, it lets me project *The Book of John Mandeville*’s Paradise references onto the Hereford world map. In the current project, I use the Hereford mappa mundi image onto which the Mandeville Paradise references are mapped. The mapping uses geometric shapes to signify the degree of certainty for my mapping. *Circles* signal Mandeville Paradise references whose name corresponds to a named point on the Hereford map. For example, both Mandeville and Hereford depict Jerusalem, and Mandeville locates in Jerusalem Christ’s bath, filled by water that flows from Paradise (see Figure 3.1). Similarly, both

27 Omeka, <http://omeka.org/about/>. On Omeka’s uses for archive building, see, in alphabetical order, French, “Introduction to Omeka”; McClurken, “Teaching and Learning with Omeka”; and Schlitz and Bodine, “The Martha Berry Digital Archive Project.”

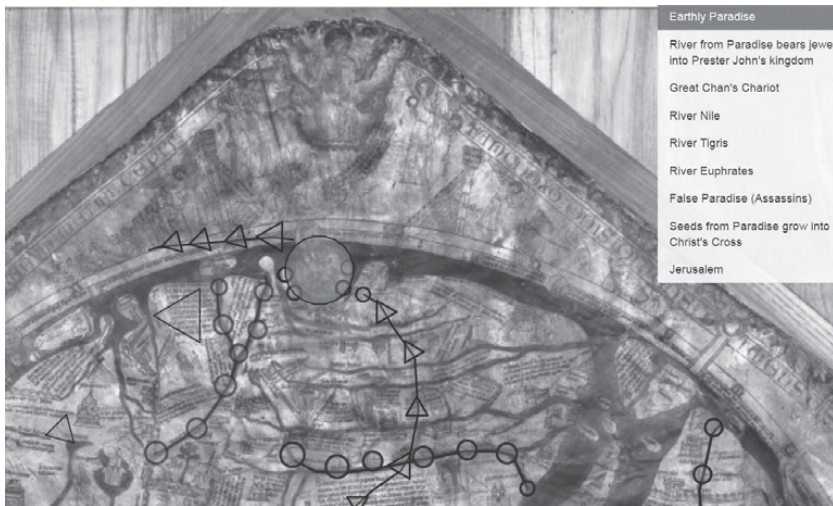
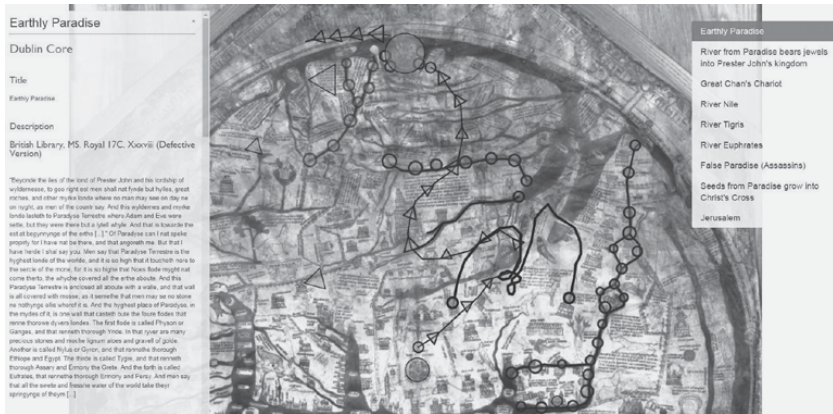
28 On the tension between interpretation and standardization—for metadata schemas in general, and for Dublin Core in particular—see Rühle, Baker, and Johnston, “User Guide.”

29 *The Book of John Mandeville*, eds. Kohanski and Benson.

30 Deluz, *Jean de Mandeville*.

31 Deluz, *Jean de Mandeville*.

32 Nowviskie, “Neatline and Visualization as Interpretation.”



Figures 3.1, 3.2. Hereford map, Earthly Paradise (covered by large circle) with rivers (indicated by small dots within the larger circle). Permission is granted by The Dean and Chapter of Hereford Cathedral and the Hereford Mappa Mundi Trust.

Mandeville and Hereford represent the Earthly Paradise enclosed within walls. The four rivers of Paradise spring from its centre, and the guarding angel drives Adam and Eve into exile (See Figure 3.2).³³

Triangles signal Mandeville references to Paradise whose location on the Hereford map I had to establish by research and some speculation: that is, by placing the

33 For the relationship between Paradise and Jerusalem, see Akbari, *Idols in the East*, 51–59; Higgins, “Defining the Earth’s Center,” 29–53; and Higgins, *Writing East*.

Mandeville references in modern geography and history, and then projecting these modern locations back onto the Hereford map. For example, the Mandeville-narrator lavishly describes Prester John's dominions and links these dominions to the Earthly Paradise through the waters of Paradise:

This emperor Prester John hath many good citees and tounes and many gret iles and large, for the lond of Ynde is departed in iles by cause of the gret floodes that cometh out of Paradis.

[This emperor Prester John has many good cities and towns and many great, large islands, for the land of India is separated into islands because of the great floods that come out of Paradise.]

(ll. 2393–96)

The Hereford map, however, does not depict the realm of Prester John, for that realm—especially at the time of the Hereford map's creation—did not have a stable location. As Christopher Taylor notes:

[Prester] John's elusiveness became a focal point of the legend, which, adapting to historical circumstance, adopted a kind of nomadic poetics. The physical location of John's kingdom was constantly (and necessarily) re-imagined in order to sustain the belief that this kingdom might actually exist, despite the failures by those who sought it. Given the existence of 'Three Indias' in the medieval period, Prester John's indeterminate locale and shifting significance allowed the European imagination to consequently imagine him *everywhere* precisely because he could neither be confirmed nor denied existence *anywhere*.³⁴

Despite the elusiveness of Prester John's realm in medieval tradition at large, the Mandeville-narrator places the realm in India. On Hereford, the realm of India takes up the top third of Asia, across the surface of the entire earth. Speculatively, I placed the reference to Prester John within this region, on the course of a river that could, potentially, be the "Sand Sea" that carries jewels and lignum aloes from the Earthly Paradise to the realm of Prester John, as other Mandeville-manuscripts (but not Royal 17 C. xxxvii) describe (see Figure 3.3).³⁵

On the Hereford map, there are three visual connections between Paradise and the world. The first is the rivers of Paradise, flowing out of Paradise and re-emerging in the world (see Figure 3.4). The second is the figures of Adam and Eve driven into exile (see Figure 3.4).

The third is the visual echo between the wall of Paradise and the wall of Jerusalem, at the world's centre; the Cross of Christ, above Jerusalem, is positioned above the figures of Adam and Eve driven from Paradise (Figure 3.5).

³⁴ Taylor, "Global Circulation as Christian Enclosure," 445–59.

³⁵ Deluz, *Jean de Mandeville*, 436.

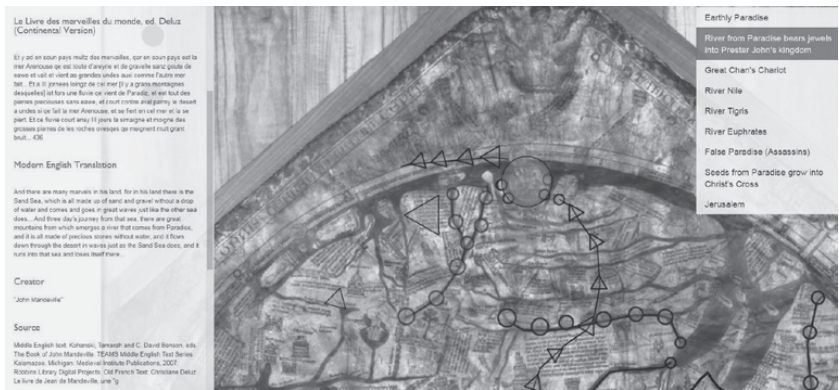


Figure 3.3. Hereford map, River from Paradise bears jewels into Prester John's kingdom (triangle). Permission is granted by The Dean and Chapter of Hereford Cathedral and the Hereford Mappa Mundi Trust.



Figure 3.4. Hereford map: Within the circle of Paradise, centre, flow its four rivers. Below Paradise on the right is the angel with the sword, driving the faint figures of Adam and Eve out of Paradise. Permission is granted by The Dean and Chapter of Hereford Cathedral and the Hereford Mappa Mundi Trust.



Figure 3.5. Hereford map: At the top, circled: Paradise; below, just above the lower border: the walled city of Jerusalem; above it, Christ's cross. Permission is granted by The Dean and Chapter of Hereford Cathedral and the Hereford Mappa Mundi Trust.

The visual connections underscore the theological connection between the Old Adam's fall and the New Adam's death and resurrection. Apart from these connections, the Hereford Paradise is an island in the Ocean, visually separated from the world by wall and water.

In comparison, the Mandeville Paradise is both separated from the world and inextricably connected to the world. Like Paradise on the Hereford map, Paradise in Mandeville is separated from the world by geographical obstacles: walls, water, hostile geography. These geographical obstacles are accentuated by the language in which they are conveyed; rhetorical enclosures mirror geographical ones. But Mandeville's Paradise is also connected to the world, far more so than Hereford's Paradise. In addition to the Paradise references it shares with Hereford (marked by circles, above), Mandeville also recalls Paradise in the episodes of the fountain of youth, the *lignum aloes* at the Great Chan's and Prester

John's courts, the jewel-bearing Sandy Sea that flows with precious stones from Paradise into Prester John's kingdom, the false Paradise of Gathalonabes and the people who feed their beloved dead to birds, so they may be carried to Paradise. All these episodes depict a Paradise that is—through distorted reflections or literal relics—deeply connected to the world; yet none of these narrative episodes has a spatial equivalent on the Hereford map. In the section that follows, I look at how the spatial poetics of Mandeville maps, and fails to map, onto the Hereford map.

The following section focuses on the Mandeville narrative episodes that have no equivalent on the Hereford map. Using Neatline to place them on the Hereford map, I analyze the spatial poetics that underlies them.

Mandeville's Spatial Poetics

The Book of John Mandeville begins its account of Paradise by doing rhetorically what Hereford does visually: that is, establishing Paradise's separation from the world. The Mandeville-narrator confesses that he can "nat speke propirly" of Paradise (l. 2705) because he does not have direct experience of it. Instead, he draws on hearsay knowledge (l. 2706). Yet this knowledge that he goes on to weave throughout his fantastic narrative amounts to an acknowledgement that Paradise is unknowable and inaccessible.

In every geographical way imaginable, Paradise is situated *beyond*. Paradise lies beyond human civilization ("beyonde the iles of the lond of Prester John and his lordship of wylderneshe," l. 2696); beyond ordinary time (it is surrounded by a "other myrke londe where no man may see on day ne on nyght, as men of the countré say," ll. 2697–8); beyond the reach of post-lapsarian biblical history ("it is so highe that Noes flode myght nat come therto, the whyche covered all the erthe aboute," ll. 2708–9), and almost beyond the mortal, sublunary world ("it is so high that it toucheth nere to the sercle of the mone," l. 2707–8). Paradise's wall—as emphatic in Mandeville verbally as on Hereford visually—emphasizes the profound inaccessibility of Paradise:

And this Paradyse Terrestre is enclosed all aboute
with a walle, and that wall is all covered with mosse,
as it semethe that men may se
no stone ne nothyngne ellis wherof it is.

[And this Earthly Paradise is enclosed all around with a wall, and that wall is all covered with moss, as it seems, so that people are not able to see any stone or anything else that it is made of.]

(ll. 2705–11)

As the narrator notes, the wall is impenetrable not just physically, but also visually and epistemologically, for it is itself enclosed in moss, and its material remains entirely mysterious. This inaccessibility of Paradise receives further emphasis at the end of the episode. Mandeville repeats formulaic declarations of inaccessibility no fewer than five times in one passage:



Figure 3.6. The Hereford map: Within the Earthly Paradise (encircled, above), the rivers stop before the wall. Outside the Earthly Paradise, the four rivers (from the top: Indus (Ganges), Tigris, Euphrates, and Nile) are not visually connected to Paradise. Permission is granted by The Dean and Chapter of Hereford Cathedral and the Hereford Mappa Mundi Trust.

And ye shall understonde that no man lyvyng may go unto that Paradyse. For by land he may nat go for wylde bestes which ar in the wyldernes, and for hylles and roches where no man may passe. Ne by those ryvers may no man passe, for they come with so greate course and so great wawes that no shyp may go ne sayle ageyne them.

[And you must understand that no living person may go into that Paradise. For by land they cannot go because of the wild beasts which are in the wilderness, and because of hills and rocks where no one can pass. Nor can any person pass by those rivers, because they come with so great a flow and such great waves that no ship is able to sail against them.]

(ll. 2729–33)

Mandeville's enumerations of barriers frame the account of Paradise, so that the barriers enclose Paradise textually as they do geographically. As Iain Macleod Higgins

notes, the account of Paradise is “framed by two paragraphs that emphasize Paradise’s apartness.”³⁶ Between them is the Mandeville-author’s description of what he calls “þe myddel of hit,” namely the four rivers that issue forth from Paradise.

In the *Book*, the four rivers of Paradise are the only things that escape Paradise’s enclosure. On the Hereford map, the four rivers of Paradise appear both within the Earthly Paradise itself and outside it. But within Paradise, the rivers stop before the wall, and the Earthly Paradise is enclosed by walls and the encircling ocean. And outside Paradise, none of Hereford’s earthly rivers has a visual connection to their source waters in Paradise.

By contrast, Mandeville stresses the physical connections between the rivers of Paradise and the world’s geography, emphasizing that in the rivers of Paradise, all the world’s fresh waters have their origin. As Paradise’s rivers traverse the earth’s geography, they scatter it with the fragments and relics of Paradise: the rivers of Paradise bring fresh water into the world, precious wood (lignum aloes) to the Great Khan’s court, jewels and wonder to Prester John’s Kingdom. Yet at the same time the rivers replenish the world with water and with wonder, they serve as yet another implacable barrier, for their currents and waves are so strong that those who attempt to sail them come to bad ends:

Many greate lordes have asayed many tymes to go by those ryvers
to Paradyse, but they myght nat spede in their wey. For som dyed for wery of
rowynge, som wex blynde, and som deefe for noyse of the waters. So no man may
passe there but thorough special grace of God.

[Many great lords have tried many times to go by those rivers to Paradise, but they were not able to succeed in their way. For some died exhausted by rowing, some went blind, and some deaf because of the noise of the waters. So no one may pass there but through special grace of God.]

(ll. 2733–36)

The anaphora (“som...som...som”) shows the temporal power of the “many greate lordes” falling apart into wretched fragments; the enumeration of their disastrous fates leads inescapably into the narrator’s reiterated assertion that “no man may passe there but thorough special grace of God.” Paradise’s fluid boundaries both link it to the world and keep it inaccessible.

Tracing Paradise and its rivers in the *Book* and on the map reveals that the Mandeville narrative is more paradoxical in its enclosure of Paradise. The Hereford map shows Paradise as an enclosed island; it does not visually link the rivers within Paradise to the same rivers flowing through the world. The *Book*, by contrast, emphasizes both Paradise’s apartness and its connectedness to the world.

The emphasis in the *Book* on Paradise’s simultaneous presence and loss is even more evident in its one story of human beings reaching Paradise—a story that is both a failure and—in the long, long run—a victory. The story appears quite early in the text, in the description of the Cross; the Mandeville-narrator claims that it is part of the Greek Orthodox Scriptures:

36 Higgins, *Writing East*, 204.

As Grekes and Cristin men that dwelleth over the see seyn, that the tre of the Cros that we callen cipresse was of the tree that Adam eet of the apple, and so fynde [fol. 9r] they writen. And they sayn that her scripture telleth that Adam was sike and bad his son Seth that he sholde go to Paradis and pray the angel that kepide Paradis that he wolde sende hym of the oyle of mercy to anynty therwith his membris that he myght have heel. And Seth wente theder, but the angel wold nought lete hym in, but seyde to hym that he myght nought have the oyle of mercy. But tho he toke hym foure graynes of the same tre that Adam eet of the apple and bad hym als so sone as his fader were deed, he sholde put tho graynes under his tonge and grave hym so. And he dide so. And of these foure graynes spronge foure trees, as the angel seyde, which sholde bere a fruyt, thorgh which fruyt Adam sholde be saved. Of which trees was maad the Cros that bare God Jhesu Crist, that sweet fruyt thorgh which Adam and alle that come of hym were saved and deliverid fram eyndelys deth, but hit be here owen defeaute.

(ll. 135–48)

[The Greeks [i.e. the Greek Orthodox Christians] and likewise the Christians who live over the sea say that the tree of the cross that we call cypress was from the apple tree whose fruit Adam tasted ... [W]hen Adam was ill, he told his son Seth to go to Paradise and beseech the angel guarding Paradise to send him some oil of mercy to anoint his limbs with it and restore his health. Seth went there, but the angel would not let him in, but told him that he could not have any oil of mercy. But then he [the angel] gave him [Seth] four seeds from the same tree that Adam ate the apple from, and told him to place them in his father's mouth as soon as his father were dead, and bury him so. And he [Seth] did this. And from the four seeds grew four trees, as the angel said, which would bear a fruit through which Adam would be saved. From these trees the cross was made that bore Jesus Christ, that sweet fruit through which Adam and those descended from him are saved and delivered from perpetual death, if it were not for their own fault.]

Higgins notes that Seth's journey to the Earthly Paradise is successful, unlike the disastrous journeys that the Mandeville-narrator mentions later in the text.³⁷ But this is a very qualified success. Seth's journey ends not in Paradise itself, but at the gate, for the angel will not let him in. Moreover, Adam's failure is more vivid and poignant than Seth's success. The banished Adam asks for the "oyle of mercy" (oil of mercy) that foreshadows the sacrament of extreme unction: what would be at once physical and spiritual restoration to his former state in Paradise. But the banishment holds. He is denied the oil, and the mercy, that he seeks. The final consequence of Adam's sin, which is physical death, must at last overtake him, and the only way back into Paradise, both for him and for his descendants, is through Christ.

The story establishes, as Higgins remarks, an organic connection between the two geographical and theological "poles" of Mandeville's itinerary, Jerusalem and the Earthly Paradise. The traditional typological connection between the Old Adam and the New is represented through a biological connection between the Tree that occasioned the

³⁷ Higgins, *Writing East*, 78–80.

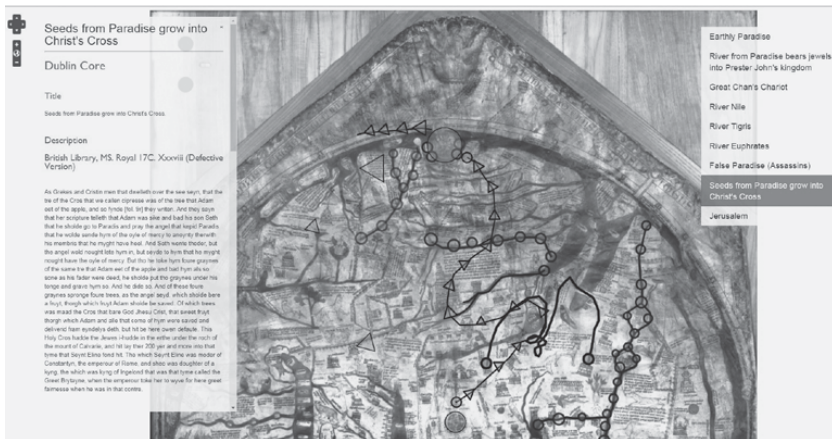


Figure 3.7. The Hereford map: The triangle-studded line from the Cross (below) to Paradise and the angel (above) to the company of the blessed in the afterlife (top left). In the bottom right hand corner, the winding line is the Hereford cartographer's own tracing of the Israelites' journey out of Egypt into the Promised Land; I have reinforced it with a thick black line and circles at beginning and end, for visibility. The triangle-studded line of Seth's journey speculates that, up to a point, Seth traces the Israelites' journey—though he does not reach *his* desired land. Permission is granted by The Dean and Chapter of Hereford Cathedral and the Hereford Mappa Mundi Trust.

fall of humankind and the Tree that occasioned its redemption. Thus the brief legend encompasses all of salvation history, from the fall of humankind, to its redemption, to a foreshadowing (in that last word, “defaute”) of its future judgment—and encompasses, at the same time, the space of the world, from Paradise to the Cross, from the Cross to Paradise. In a similar way, the Hereford mappa mundi displays—at its very top—the Last Judgment; just below it, Adam and Eve’s banishment from Paradise; and just below that, the Crucifixion. Here, too, salvation history frames the world’s space and time.

But despite these commonalities, mapping the Seth-episode against the Hereford map was the most extensively speculative. Since the trees of the Cross sprang from the seeds of Paradise buried under the dead Adam's tongue, I speculated that Adam would have been buried near Golgotha—and so Seth would have set out from there, too. I then traced Seth's journey towards Paradise in a fairly straight line, leading Seth to the encounter with the Hereford angel pictured below the Earthly Paradise—and also reaching beyond the map among the blessed souls at the Last Judgment, to indicate the episode's reach into eternity. I further assumed that Seth would have returned the way he came and did not add a different trajectory for his return. As I drew the line on the Hereford map, each point was an interpretive decision.

The point-by-point close reading of the Hereford map revealed a feature I had not noticed before: another journey marked as a winding line on the map—that of the Israelites travelling out of Egypt to the Promised Land. Even though my line tracing Seth's journey to Paradise is fabrication, it is fabrication that fits with the spatial poetics of the map—a map that visualizes not only places but also their history; not only space but also time.

In *The Book of John Mandeville*, the Seth-episode is one of numerous fragments of Paradise in Mandeville's narrative geography. Mapping the episode onto the Hereford map, drawing its trajectory from Jerusalem to Paradise and into the timeless life of Christ beyond, gives visual form to the way that, in the *Book*, Paradise pervades the world's geography. In Hereford, Paradise is in the margins; in Mandeville, Paradise surfaces and resurfaces throughout the world.

Mapping Mandeville against Hereford examines geographical representations of Paradise by bringing two major texts together, the most detailed surviving mappa mundi and the most popular travel narrative of the Middle Ages. Reading Mandeville against the Hereford map clarifies the thematics of Mandeville's Paradise references. Unlike the Hereford mappa mundi, where Paradise is contained within its walls in the world's uttermost East, the *Book* is haunted by Paradise's insistent presence and just as insistent absence. What Mandeville's references to Paradise have in common—especially those that cannot be mapped onto the Hereford mappa mundi—is fragmentation. Whether they are Paradise's rivers branching out into smaller and smaller distributaries, or the relics they bear into the world, or the seeds from the forbidden tree, Paradise enters the world in fragments. The text is thick with aspirations towards Paradise and attempts to reach it in this life, by unaided human effort; but time after time the attempts to grasp Paradise fail.

Significantly, the narrative about Seth is the only episode in Mandeville where Paradise enters the world not simply as fragments, but as fragments that grow into a new whole: the initial relic of Paradise (the four seeds) grows into four trees, which then grow into a cross, which in turn "bears good fruit, Jesus Christ." Though Paradise persists in the world, the recovery of Paradise is only possible through Christ's costly grace and a passage through death. Project Paradise, then, is in proper Mandevillian fashion a misnomer: its insights emerged not from mapping Paradise, but from failing to map it; not, ultimately, from plotting journeys to Paradise onto virtual space, but from gathering up their fragments and shoring them up against our ruins—virtually reconstructing literary worlds by re-engaging in their spatial storytelling practices.

Conclusion

As noted earlier, the methodological basis of this study is the notion of deformation: a reading strategy that invades a literary text and changes its shape in order to defamiliarize the text and reveal new interpretive possibilities. Project Paradise practices deformation by pulling the *Book's* references to Paradise out of their immediate narrative context and projecting them on the map. But its deformation, as the portmanteau term suggests, is not just about changing the shape of a text but about performing one of its aspects. In this study, the project de-contextualizes the Mandeville passages from their immediate narrative, but it also restores a wider historical context, using the Mandeville passages to re-enact medieval spatial storytelling practices. For Neatline and medieval world maps are interestingly akin to each other: both are platforms that transform space into time, time into space.

Medieval world maps do not perform objective mapping—the georectifiable representation of geographical space—but narrative mapping—the layering of time, theology, history, the story of the world, onto geographic space. The Hereford map, for instance, has nearly 1,100 map legends, or inscriptions. Most of these are names of geographical features; but others contain “detailed cosmological, ethnographical, historical, theological, and zoological information (or at least lore).”³⁸ Not only are these mini-narratives embedded in the map’s space, but narrative actually shapes space, as is the case with the Hereford map’s depiction of Jerusalem and the Holy Land:

On the Hereford map this area gets disproportionate space not only because there was a lot to show but also to emphasize its importance. Less well-known or interesting regions were correspondingly shrunk in size. Elsewhere, places were sometimes dislocated to make room for pictures or lengthy descriptive legends.³⁹

Thus space expands with the accumulation of history; the map gives visual, spatial form to narrative. Moreover, studies of the Hereford map’s making suggest that narrative, textual sources were at least as influential as visual sources in the map’s construction. The *Expositio mappe mundi*, extant in two fifteenth-century manuscripts but likely dating from the twelfth century, is a detailed textual map that not only describes the world and its places, but describes how they can be assembled spatially, into a visual map.⁴⁰ Scott Westrem notes,

If it is not the recipe for making a mappamundi, EMM is certainly a careful record of the content of an existing one (so careful, in fact, that even if it was originally composed only as a *descriptio*, it could have been used to produce another). The text is in essence a collection of 484 data items, which range from detailed reports about Asian locations to simple toponyms listed according to their location along coastlines or rivers. ... That cartographers did indeed employ EMM (or a text very like it) is proved by the inscriptions on the Hereford Map, where more than 400 of these 484 existing data items appear verbatim or nearly so.⁴¹

While discrepancies exist between EMM and Hereford, the many word-for-word textual parallels between the two maps strongly suggest that a text very like EMM might have been instrumental in the making of Hereford. More broadly, EMM also attests to a culture of cartographic thought and practice that navigated seamlessly between the textual and the spatial, between description and map, story and space.⁴²

Like the medieval world maps of which Hereford is the most detailed surviving example, Neatline is a platform for telling stories on maps, through time and archives. As Bethany Nowviskie notes,

38 Westrem, “Making a Mappamundi,” 19–33.

39 Edson, *World Map*, 18.

40 Dalché, “Décrire le monde.” Cited in Westrem, “Making a Mappamundi,” 19–33n14.

41 Westrem, “Making a Mappamundi,” 28.

42 Westrem, “Making a Mappamundi.”

Neatline sees visualization itself as part of the interpretive process of humanities scholarship—not as an algorithmically-generated, push-button result or a macro-view for distant reading—but as something created minutely, manually, and iteratively, to draw our attention to small things and unfold it there.⁴³

Digital mapping applications like ArcGIS or Carto draw their display from an underlying database of georeferenced data. That is, the user uploads a file with location codes and geographical longitude and latitude; the software automatically translates this file into visual points on the map. By contrast, in Neatline, the data is not batch-ingested and automatically made visible at coordinates derived from the data. Rather, each point placed on the map is an individual act of interpretation: the user must choose the location on the map, choose the Omeka item, and create the visual representation on the map. It is the user's choice and physical action, and not previously stored measurements, that decide the visual shape of the map.

The affordances of the Neatline platform further privilege contextualized close reading. Each item has metadata associated with it—metadata, especially in this project, that is based on a reading and interpretation of the text. So each point and its metadata lead further in, deeper into the literary text—into a reference to Paradise mapped against the wider Mandeville tradition as well as against the Hereford map. And each item, through tagging and through Neatline's styling options, allows users to visually encode levels of certainty about my mapping: circles for certain matches between map and text, triangles for speculative ones, found by stepping out of the text, off the map, into the wider network of medieval traditions.

The work of constructing the map begins in traditional philology, with comparing manuscript variants, tracing the thematics of a literary tradition, deciphering and correlating the narratives in map and codex. But as the Mandeville narrator draws nearer the Earthly Paradise, projecting the narrative onto the map begins to resemble the Mandeville narrator's own synthetic approach. The geo-temporal exhibit is itself a narrative of space, seeking to reconcile—sometimes speculatively—several related traditions: Mandeville, Hereford, and the encyclopaedic traditions on which both works draw. The Mandeville narrator, for instance, synthesizes and adapts the accounts of his sources—William of Boldensele, (Pseudo-) William of Tripoli, John le Long's version of Odoric's *Relatio*—in his geography of the East.⁴⁴ Similarly, I synthesize the Hereford map and the *Book* to generate the visual synthesis of the annotated Hereford map. The close reading and construction work enabled by Neatline allows me to re-perform the synthetic approach, with liberal interposition of his own subjectivity, through which the Mandeville narrator weaves together travel narrative and encyclopaedic tradition to build the story-space of his invented travels.

The construction of the digital map becomes a hybrid between traditional scholarly practice and a more performative approach to the materials. In the geo-temporal

⁴³ Nowviskie, "Neatline and Visualization as Interpretation."

⁴⁴ Higgins, *Writing East*.

archive that emerges, the shape of Mandeville's own narrative is hidden; quotations are pulled out of their immediate contexts and pinned onto the map. But that very action places them in a wider context—that of the thick cultural landscape from which the Mandeville multitext grows, a landscape formed by the mappae mundi and by the encyclopaedic, biblical, and historical traditions that the mappae mundi give spatial shape to.⁴⁵ Likewise, the digital exhibit gives a spatial shape to the performance of reading the Mandeville-narrator's approach to Paradise. That is, to uncover the Mandeville passages about Paradise, readers must advance not from one leaf of a codex to another, but from point to spatial point on the digital Hereford map. Thus, Neatline's interface requires both builders and users of the exhibit to navigate between traditions, between space and text, across intersections between map and narrative.

On the Hereford mappa mundi, the digital exhibit maps, point by uncertain point, the narrative traces of an absent Paradise. The map and the book touch but do not overlap. The digital exhibit is fragmentary and imperfect. It is as yet limited to a single theme, a single instance of the Mandeville multi-text. It provides not a seamless experience of the text, but one interrupted by layers of metadata that document the exhibit's interpretive decisions. But like the seeds of knowledge that Seth carries back from Paradise, the exhibit's fragmentary nature bears the "good fruit" of insight. Point by point and click by click, the builder and readers of the exhibit are led through the palpable frictions and uncertainties of projecting the book onto the map, of synthesizing medieval traditions in a digital medium, of transforming (as the Hereford cartographer and the Mandeville narrator both do) story into space and space into story. More than either a print edition of the text or a printed edition of the map, Neatline facilitates the interplay between text and map, between visual and textual narrative. The digital platform enables an experiential understanding of medieval modes of thinking about space and its representations in map and story. And in the process, it reveals the *Book's* own spatial poetics, its portrayal of a Paradise at once absent and present, mapped yet unknowable, walled off from the world yet inextricably woven into its fabric.

Bibliography

- Akbari, Suzanne Conklin. *Idols in the East*. Ithaca: Cornell University Press, 2009.
- Albu, Emily. *The Medieval Peutinger Map: Imperial Roman Revival in a German Empire*. Cambridge: Cambridge University Press, 2014.
- Bodenhamer, David J., John Corrigan, and Trevor M. Harris. *The Spatial Humanities: GIS and the Future of Humanities Scholarship*. Bloomington: Indiana University Press, 2010.

⁴⁵ Neatline also offers other mechanisms—such as its Timeline and Waypoints—that could foreground the Mandeville text rather than the map. As Matthew Evan Davis notes (private communication, October 2016), this change of perspective could even more richly contextualize the Mandeville narrative by embedding it into the literal cultural landscape of the map.

- Bolintineanu, Alexandra. "Towards A Poetics of Marvellous Spaces in Old and Middle English Narrative." PhD dissertation, University of Toronto, 2012.
- Bonnell, Jennifer and Marcel Fortin, eds. *Historical GIS Research in Canada*. Calgary: University of Calgary Press, 2013.
- Dalché, Patrick Gauthier. "Décrire le monde et situer les lieux au XIIe siècle: L'Expositio mappe mundi et la généalogie de la mappemonde de Hereford." *Mélanges de l'École française de Rome. Antiquité-Moyen Âge* 112.1. Rome: École française de Rome, forthcoming.
- DECIMA website <http://decima-map.net/a-new-way-to-study-florence/>
- Deluz, Christiane. *Jean de Mandeville: Le Livre des Merveilles du Monde*. Paris: CNRS Editions, 2000.
- . *Le livre de Jehan de Mandeville: Une "géographie" au XIVe siècle*. Louvain-la-Neuve: Institut d'Études Médiévales de l'Université Catholique de Louvain, 1988.
- Drucker, Johanna, and Bethany Nowviskie. "Speculative Computing: Aesthetic Provocations in Humanities Computing." In *A Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens, and John Unsworth, 431–47. Malden: Blackwell, 2004.
- Edson, Evelyn. *Mapping Time and Space: How Medieval Mapmakers Viewed Their World*. London: British Library, 1997.
- . *The World Map, 1300–1492: The Persistence of Tradition and Transformation*. Baltimore: Johns Hopkins University Press, 2007.
- Foys, Martin, and Shannon Bradshaw. "Developing Digital Mappaemundi: An Agile Mode for Annotating Medieval Maps." *Digital Medievalist* 7 (2011). www.digitalmedievalist.org/journal/7/foys/.
- French, Amanda. "Introduction to Omeka." <http://amandafrench.net/2013/11/12/introduction-to-omeka-lesson-plan/>.
- Friedman, John B. "Cultural Conflicts in Medieval Maps." In *Implicit Understandings: Observing, Reporting and Reflecting on the Encounters between Europeans and Other Peoples in the Early Modern Era*, edited by Stuart B. Schwartz, 64–95. Cambridge: Cambridge University Press, 1994.
- Greenlee, John Wyatt. "The Mapping Mandeville Project." <http://historiacartarum.org/john-mandeville-and-the-hereford-map-2/>.
- Gregory, Ian and Alistair Geddes. *Toward Spatial Humanities: Historical GIS and Spatial History*. Bloomington: Indiana University Press, 2014.
- Higgins, Iain MacLeod. "Defining the Earth's Center in a Medieval 'Multi-text': Jerusalem in *The Book of John Mandeville*." In *Text and Territory. Geographical Imagination in the European Middle Ages*, edited by Sylvia Tomasch and Sealy Gilles, 29–53. Philadelphia: University of Philadelphia Press, 1998.
- . *Writing East. The "Travels" of Sir John Mandeville*. Pennsylvania: University of Pennsylvania Press, 1997.
- Hillier, Amy, and Anne Kelly Knowles. *Placing History: How Maps, Spatial Data, and GIS Are Changing Historical Scholarship*. Redlands: Esri Press, 2008.

- Jenstad, Janelle, ed. *MoEML: The Map of Early Modern London*. 2012. <https://mapoflondon.uvic.ca/>.
- Kohanski, Tamarah, ed. *The Book of John Mandeville: An Edition of the Pynson Text*. Tempe: Arizona Center for Medieval and Renaissance Studies, 2001.
- Kohanski, Tamarah, and C. David Benson, eds. *The Book of John Mandeville*, Kalamazoo: Medieval Institute Publications, 2007.
- McClurken, Jeffrey W. "Teaching and Learning with Omeka: Discomfort, Play, and Creating Public, Online, Digital Collections." Omeka. <http://omeka.org/about/>.
- Montello, Daniel R., Karl Grossner, and Donald G. Janelle. *Space in Mind: Concepts for Spatial Learning and Education*, Cambridge, MA: The MIT Press, 2014.
- Moseley, Charles W. R. D. "Behaim's Globe and 'Mandeville's Travels.'" *Imago Mundi* 33 (1981): 89–91.
- Nowviskie, Bethany. "Neatline and visualization as interpretation" (blog), November 2, 2014. <http://nowviskie.org/2014/neatline-and-visualization-as-interpretation/>.
- Project Paradise. <http://medievalparadise.library.utoronto.ca/neatline/show/project-paradise>.
- Rühle, Stephanie, Tom Baker, and Pete Johnston. "User Guide," last modified September 6, 2011. http://wiki.dublincore.org/index.php/User_Guide.
- Samuels, Lisa, and Jerome McGann. "Deformance and Interpretation." *New Literary History* 30.1 (1999): 25–56.
- Saunders, Corinne J. *The Forest of Medieval Romance: Avernus, Broceliande, Arden*. Cambridge: Brewer, 1993.
- Scafi, Alessandro. *Mapping Paradise: A History of Heaven on Earth*. Chicago: University of Chicago Press, 2006.
- Schlitz, Stephanie, and Garrick S. Bodine, "The Martha Berry Digital Archive Project: A Case Study in Experimental pEDagogy." In *Code4Lib Journal* (2012). <http://journal.code4lib.org/articles/6823>.
- Seymour, Michael C. "The Origin of the Egerton Version of Mandeville's Travels." *Medium Aevum* 30 (1961): 159–69.
- Taylor, Christopher. "Global Circulation as Christian Enclosure: Legend, Empire, and the Nomadic Prester John." *Literature Compass* 11 (2014): 445–59.
- Tomasch, Sylvia. "Reviews: *The Hereford Map*, Scott D. Westrem," *Speculum* 78 (2003): 921–23.
- Westrem, Scott D. *The Hereford Map: A Transcription and Translation of the Legends with Commentary*. Turnhout: Brepols, 2001.
- . "Making a Mappamundi: The Hereford Map." *Terrae Incognitæ* 34 (2002): 19–33.
- Woodward, David. "Medieval Mappaemundi." In *The History of Cartography, I: Cartography in Prehistoric, Ancient, and Medieval Europe and the Mediterranean*, edited by J. B. Harley and David Woodward, 286–370. Chicago: University of Chicago Press, 1987.
- . "Reality, Symbolism, Time, and Space in Medieval World Maps." *Annals of the Association of American Geographers* 75.4 (1985): 510–21.

Wrisley, David J. "Spatial Humanities: An Agenda for Pre-Modern Research." *Porphyra* 22 (2014): 96–107.

———. "'Locating Medieval French,' or Why We Collect and Visualize the Geographic Information of Texts." Special issue, *Speculum* 91 (October 2017): 145–69.

Yeager, Suzanne M. *Jerusalem in Medieval Narrative*. Cambridge: Cambridge University Press, 2008.

Zumthor, Paul. *La mesure du Monde: Représentation de l'Espace au Moyen Âge*. Paris: Seuil, 1993.

Chapter 4

GHASTLY VIGNETTES: PIERCE THE PLOUGHMAN'S CREDE, THE GHOST OF SHAKESPEARE'S BLACKFRIARS, AND THE FUTURE OF THE DIGITAL PAST

JIM KNOWLES

"The past we reconstruct will shape the future we must live."

—Lee Patterson, "Historical Criticism and the Claims of Humanism"¹

"Langland would never have made such a gaffe."

—George Kane, "Some Fourteenth-Century 'Political' Poems"²

It is a critical commonplace in Middle English studies to situate *Pierce the Ploughman's Crede* in the urban streetscape of late fourteenth-century London. This placement is justified, if mentioned at all, by the claim that certain details in the poem's rich architectural description of the London Blackfriars church and priory are matched by what is known of the actual site of medieval Blackfriars. But when the trail of scholarship is traced back to its source, it becomes clear that the evidence for the link between *Crede* and London Blackfriars, and thus also the link between *Crede* and London more generally, is very scant. After casting sufficient doubt on the accuracy of the London Blackfriars connection, this essay sets out to describe and contextualize, as counter-example, the case for Oxford as a site at least as likely as London as the setting for *Pierce the Ploughman's Crede*. In doing so, it also raises some higher-level historiographical questions about the relation between literary "description" and real material stuff on the ground. The method of investigation in the Oxford case study, moreover, is aided by, or, more accurately, shaped by, the use of new digital tools for historical visualization, specifically by the author's creation (with Michal Koszycki) of three-dimensional digital models of the lost Franciscan and Dominican convents in medieval Oxford. Thus this study, in its methods, results, and implications, falls under the category of the digital humanities, and the essay will be punctuated with several brief asides on the value and limitations of these technical innovations on the field of medieval studies. Are we at a moment when new technologies force us to look again at the vexed relationship between text and context that historicist literary criticism has struggled to theorize for decades? Do the tools provided by the digital humanities give us some new methodological purchase on these old problems? The essay argues,

¹ Patterson, "Historical Criticism," 45.

² Kane, "Some Fourteenth-Century 'Political' Poems," 89.

obliquely, that the answer is both yes and no. New tools give us new perspectives on old problems, but run the risk of replacing fully contextualized analysis with mere “data.”

The Ghost of Shakespeare’s Blackfriars

First, the history of an error. It has not always been the case that scholars interested in the anonymous fourteenth-century English poem known as *Pierce the Ploughman’s Crede* (ca. 1394) have thought it necessary to identify the location of the buildings “described” by the poem’s narrator.³ Thomas Warton, writing his massive *History of English Poetry* in the 1770s, introduces his discussion of *Crede* with a long and even-handed account of the friars’ strengths and weaknesses, including their supposed penchant for architectural excess.⁴ But Warton, who is writing from Oxford, and whose introduction to the poem mentions the opulent libraries of the Franciscans in both London and Oxford, takes no position on whether or not the buildings described in the poem are in London, Oxford, or elsewhere. The question seems not to have arisen.

A century later, when the Reverend Walter Skeat published his first edition of the poem for the Early English Text Society, he does not venture a guess about the location of the mendicant foundations visited by the poem’s narrator. In his notes to relevant passages, Skeat instead quotes a Wycliffite text on the general building practices of the friars, and the scriptural allusions seem designed precisely to universalize the condemnation, not to localize it:

Swilk maner of men bigging thus biggings semen to turn bred into stones; that is to sey, the bred of the pore, that is, almis beggid, into hepis of stonis, that is, into stonen howsis costly and superflew, and therfor they semen werrar than the fend, that askid stonis into bred.⁵

[thus the men building such buildings seem to turn bread into stones; that is to say, the bread of the poor, which is alms, into heaps of stones, that is, into costly and superfluous stone houses, and therefore they seem worse than the devil, who turned stones into bread.]

3 In quoting from or referring to the text of *Pierce the Ploughman’s Crede*, I use the edition by Helen Barr in *The Piers Plowman Tradition*. Line numbers are given parenthetically in the text. Although the poems appear together in several fifteenth-century manuscripts and are included together in Owen Rogers’s 1561 print edition of *Piers Plowman*, the poem called *Pierce the Ploughman’s Crede*, written by an anonymous Wycliffite poet, and the poem called *Piers Plowman*, attributed to William Langland (in three versions A, B, and C), are almost certainly not by the same author. See Barr, *Piers Plowman Tradition*, introduction, xi.

4 Warton, *A History of English Poetry*. Warton’s discussion of *Crede* and lengthy quotation of the poem, including the Blackfriars passage, is in vol. 1, section 9.

5 Skeat, *Pierce the Ploughmans crede*, 49; quoting Todd, *An Apology for Lollard Doctrines attributed to Wycliffe*.

Nowhere in the preface or the notes of Skeat's 1867 edition does he attempt to link *Pierce the Ploughman's Crede* to a specific location. But when Skeat revised his edition in 1906, he added a few sentences claiming the poem for London (on dialect grounds) and adding that the architectural tour of the four orders' buildings would fit nicely with the geographical disposition of the known sites in medieval London.

In the same decade that Skeat proposed for the first time that the narrator's itinerary in *Crede* might be fixed to a specific urban streetscape, the English architectural historian Sir Alfred Clapham was at work on the first attempt since the forced closure of London's theatres during the Civil War (1640s) to identify the exact location of Shakespeare's indoor theatre—named Blackfriars because it was built within the precinct of the pre-Reformation Dominican priory at Ludgate. Clapham's conjectural reconstruction of the ground plan of London Blackfriars was first published in the journal *Archaeologia* in 1912, and reprinted in 1913 in his book *Some Famous Buildings and Their Story*, co-authored with Walter Godfrey.⁶ Along with the ground plan, Clapham and Godfrey printed a “free rendering” in modern English of the passage from *Pierce the Ploughman's Crede*, which they took to refer “almost certainly to the London house of the Blackfriars, since, whenever the details can be tested, they are found to be accurate.” Their account of the poem's documentary value goes on: “the lines are not only of peculiar interest from their remarkably vivid *description* of a great friary in its architectural prime, but also from their containing the *most graphic purely architectural description* in the whole range of English medieval literature.”⁷ Clapham may have been right about this being the “most graphic” text of its kind in Middle English (I have no claim to stake there). Rather, what is important for my account of what follows in the scholarly reception of the *Crede* is Clapham's use of the word “description” here (twice) as an appropriate label to be applied to a passage of poetry. This way of thinking about what this poetry *does* will have long-range consequences, as we shall see.

Before continuing that history, however, it will be helpful to have some of this “graphic description” in front of us. In the course of the poem the narrator tells of his encounter with each of the four orders of friars—the Franciscans, Dominicans, Augustinians, and Carmelites—whose members he has set out to consult about the trouble he is having in learning his creed (line 6). He visits, on foot, each of the friaries in turn. As he leaves the Franciscan precinct and presses on “to the prechoures” (Dominicans, or Blackfriars), he remarks

Swich a bild bold, y-buld opon erthe heighte
 Say I nought in certeine siththe a longe tyme.
 [Such a boldly built pile on earthly heights
 Certain, I have not seen for a long time.]⁸

6 Clapham, “On the Topography of the Dominican Priory of London”; revised and expanded in Clapham and Godfrey, *Some Famous Buildings and Their Story*; the conjectured plan is on p. 270.

7 Clapham and Godfrey, *Some Famous Buildings and Their Story*, 256 (my italics).

8 I provide Clapham's modernized English version in brackets; *Some Famous Buildings*, 256–57.

And continues

Y yemed vpon that house and yerne theron loked,
 Whough the pileres weren y-peynt and pulched ful clene,
 And queynteli i-coruen with curiouse knottes,
 With wyndowes well y-wrought wide vp o-lofte
 And thanne y entrid in and even-forth went,
 And all was walled that wone though it wid were,
 With posternes in pryuytie to pasen when hem liste;
 Orcheyardes and erberes euesed well clene,
 [And] a curious cros craftly entayled,
 With tabernacles y-tight to toten all abouten.
 [I thought on that house and long thereon looked
 How the pillars were painted and finely adorned,
 And quaintly were carven with curious knots,
 With windows well wrought, lofty and wide.
 Then entered I in and went all about,
 And everywhere there were walls around the dwelling
 With posterns for passing privately whenever they list,
 Orchards and arbours arranged with neatness,
 And a curious cross craftily constructed
 With tabernacles encircled, facing all sides]

(156–68)

After remarking on the magnificence of the “mynstre” (church) with its “[t]ombes opon tabernacles tyld opon lofte” [tombs upon tabernacles raised up on high] (181), the narrator’s tour of the precinct takes in the cloister, with its “kundites of clene tyn” and “lauoures of latun” (tin conduits and brass washbasins; 195–6); followed by the chapter house “wrought as a greet chirche” (199); and eventually the refectory (“fraytour”), which he remembers as

An halle for an heygh kinge an housholde to holden,
 With brode bordes aboute y-benched wel clene,
 With windowes of glas wrought as a chirche.

[A hall for a great king, to hold his household,
 With broad tables and benches beautifully furnished,
 And windows of glass, wrought like a church.]

(204–6)

It was this building, the frater, that would become associated with the structure referred to in sixteenth-century documents as the “upper frater.” Clapham was certain that this was the very same building that “James Burbidge converted into the celebrated Blackfriars Theatre” in 1596, and which Shakespeare’s company, the King’s Men, occupied from 1608 until Shakespeare’s death in 1616, and continued to occupy for another three decades.⁹

⁹ Clapham and Godfrey, *Some Famous Buildings*, 262; see Gurr, “London’s Blackfriars Playhouse and the Chamberlain’s Men.”

For Clapham, the Shakespeare connection was of historical interest to the nation, but it was not his main concern. Two American Shakespeareans, however, would soon drop Clapham's "almost" from "almost certainly" in an attempt to solidify the connection between Shakespeare's indoor theatre and the medieval home of the London Dominicans. In doing so, they also (incorrectly) solidified the connection between the site of London Blackfriars and its "description" in *Pierce the Ploughman's Crede*. In 1917, Joseph Quincy Adams, who would later become the first director of the Folger Shakespeare Library, published an article in *Studies in Philology* in which he proposed a wholly new ground plan for the Blackfriars precinct. He quotes the passages of *Crede* pertaining to the Blackfriars buildings, introducing them only by saying "Apparently we have a description of the Blackfriars Priory in *Pierce the Ploughman's Crede*, written about 1394."¹⁰

When another American Shakespearean, Irwin Smith, set out to compose the definitive history of the site in *Shakespeare's Blackfriars Playhouse*, published in 1964, he builds on the work of both Clapham and Adams but insists that both of the earlier accounts are incorrect due to their authors' incomplete access to, or faulty interpretation of, a cache of relevant historical documents.¹¹ Forty-six of these documents are reprinted in Smith's appendix, carefully organized as an evidentiary dossier, numbered, dated, and labelled as "Document 1," "Document 2," and so on. They include chancery court records, land transfers, surveys, bills of sale, and royal charters. "Document 44" in this list, however, is a document of a different type. It is labelled "The Blackfriars Passage in Piers the Plowman's Creed, Circa 1394."¹²

The legalistic language that undergirds Smith's method is made explicit in his introductory chapter called "Materials for Reconstruction of the Precinct." Here he writes of the Blackfriars site that "[t]he friary ... has totally vanished; it ... must therefore be reconstructed in its essentials. For this purpose four categories of *evidence* are available."¹³ He then outlines the four types as:

1. "architectural remains of abbeys and priories as revealing the usual practice of monks and friars in general, and of the Black Friars in particular, in planning their convents;
2. "traces of the actual Ludgate buildings as they have been uncovered by excavation or as they are perpetuated in modern building lines;
3. "surveys and other real estate records relating to the London precinct; and
4. "the description of the friary in *Piers the Plowman's Creed*."¹⁴

About the fourth category, Smith comments that "the *Piers* passage gives vivid details but no general plan," and concludes by suggesting that "all these groups of *data* become comprehensible only when interpreted by the system of cloistral arrangement originally

¹⁰ Adams, "The Conventual Buildings of Blackfriars, London, and the Playhouses Constructed Therein," 75.

¹¹ Smith, *Shakespeare's Blackfriars Playhouse*.

¹² Smith, *Shakespeare's Blackfriars Playhouse*, 547.

¹³ Smith, *Shakespeare's Blackfriars Playhouse*, 32 (my italics).

¹⁴ Smith, *Shakespeare's Blackfriars Playhouse*.

devised by the monastic orders and subsequently adopted, with minor modifications, by the mendicant.”¹⁵ I find Smith’s classification of this diversity of sources under the heading of “data” to be problematic (and I am very aware here that my professional commitment to literature as an object of knowledge distinct from documentary “data” is charged with bias).

As he goes on to make his case for the *Crede* fragment as admissible evidence, Smith’s language is laced with qualifications. “Nothing in the poem, to be sure,” he writes, “says specifically that the narrative relates to the Preachers’ London house in particular, but its relevancy to the Ludgate priory is all but conclusive.”¹⁶ He quotes Clapham (1913) in support of this claim, calling him a “distinguished authority on historical monuments in England,” but without drawing attention to Clapham’s own qualified ambivalence about the link between the poem and London.¹⁷ Smith goes on to say that “the identity of the *Crede*’s house with the London priory of the Friars Preachers has been accepted as probable or certain by all those who have expressed themselves on the subject,” and that “*even though no direct evidence* testifies to the splendor of the Ludgate priory in particular, the indirect evidence is strong enough to prove the case.”¹⁸ Finally, Smith winds up his presentation of the “evidence” with a critical judgment of the *Crede* that makes his posture toward the text *qua* text unambiguously clear: “[The poem’s] *descriptions*,” Smith writes, “bear evidence of direct observation, and bear none of having been embellished imaginatively for satirical purposes.”¹⁹ To my ear, there is a shift here in Smith’s epistemological position vis-à-vis his object of study. The connections he had itemized between the poem’s picture of an (unnamed) medieval Dominican friary and the London site of Shakespeare’s Blackfriars theatre had been caged around with markers of his uncertainty—to be sure, even though, all but conclusive—all perfectly ordinary qualifications for a case built on circumstantial evidence. But he seems much more certain about the poem’s status as bearer of extra-documentary (literary?) value: he claims it has none, that it fails to rise even to the level of imaginative embellishment required for satire. It fails as satire, it fails as allegory, it fails as art. But it works just fine as *data*, as *evidence*.

Why does this matter? Because it taps into a recent argument about this text and its contexts that has implications for literary history in this period going beyond the question of the location of the *Crede*’s priories. It gets to the heart of questions about whether or not polemical writing like the Wycliffite *Crede* can ever, even in its most descriptive moments, cross over the boundary of its polemically held positions to achieve artistic success on its own terms. And if Irwin Smith’s appropriation of the poem as evidence can be excused on the grounds that he never claimed to be reading the poem *qua* poem, can the same excuse

15 Smith, *Shakespeare’s Blackfriars Playhouse*, 32 (my italics).

16 Smith, *Shakespeare’s Blackfriars Playhouse*, 52.

17 Smith, *Shakespeare’s Blackfriars Playhouse*.

18 Smith, *Shakespeare’s Blackfriars Playhouse*, 53.

19 Smith, *Shakespeare’s Blackfriars Playhouse*, 54.

cover literary critics in the *Crede*'s home field of Middle English studies? Critics who ought to be prepared to interpret the poem on its own terms, in its own contexts?

In his response to James Simpson's volume in the Oxford English Literary History, Bruce Holsinger singles out exactly the same passage in the *Crede* that Irwin Smith had seized upon in gathering his body of evidence for reconstructing Shakespeare's Blackfriars. But Holsinger's reading of the Blackfriars "description" in *Pierce the Plowman's Crede* leads him to the opposite conclusion concerning its extra-documentary value.²⁰ For Holsinger, the formal and aesthetic mastery on display in this piece of "Lollard ekphrasis" is indicative of what he calls a "situated aesthetics: the self-conscious, particular manipulation of conventional literary form and style (not theme or argument alone) to the advancement of even the most radical social and political agendas."²¹ Contra Simpson—who subsumes the *Crede*, along with all Wycliffite literary productions, under the aegis of a "revolutionary" mode of writing whose inferiority to the "reformist" mode (exemplified by William Langland's *Piers Plowman*) is marked by its insistent narrowing of figural uncertainties to literal certainties—Holsinger finds in the poem a "careful, deliberate, and self-conscious literary craftsmanship" that is inseparable from the polemical effect of its antifraternal critique.²² In this respect, Holsinger argues, "it matters"

and matters profoundly that *Pierce the Plowman's Crede* is a poem: an ekphrastic poem able to beguile its readers with the aesthetic mimesis of poetry—prosody, diction, alliteration, metaphor, syntax—while using this beguilement to instill a critical awareness of clerical wealth and the idolatry of visual culture.²³

Nowhere does Holsinger suggest that this "ekphrastic display"²⁴ must be understood as a literal description, in the manner of a tourist's guidebook, of a particular building or set of buildings. To be properly ekphrastic, rather, the poem must resist such a literalist reading.

Likewise, if Holsinger's ekphrastic reading of the Blackfriars passage is right (as I think it is), we ought not to expect this poem to deliver exact correspondences between the linguistic signs of the text and their material brick-and-mortar referents. Why should we? But we do. Or rather we have done, traditionally, habitually, in making sense of this poem. In citing Ian Doyle's 1959 discovery of an unnoticed copy of the Blackfriars fragment in a British Library manuscript (MS Harley 78), Holsinger inadvertently brings full circle the scholarly history of the passage and its reception. "The fragment of the poem surviving in the Harley manuscript," he notes,

represents the bulk of the section on the Dominican house, which may have been lifted out of the larger work for its architectural verisimilitude to the

²⁰ Holsinger, "Lollard Ekphrasis," in a special issue of *Journal of Medieval and Early Modern Studies* devoted to responses to Simpson's *Reform and Cultural Revolution*.

²¹ Holsinger, "Lollard Ekphrasis," 74.

²² Holsinger, "Lollard Ekphrasis," 74.

²³ Holsinger, "Lollard Ekphrasis," 83.

²⁴ Holsinger, "Lollard Ekphrasis," 84.

London house of the Blackfriars: the poem's descriptions of the conventual church, the chapter house, the cloister, and the refectory all correspond in important details (or so Doyle contended) with the historical building as the poet probably knew it.²⁵

When we track down the basis for Doyle's contention, however, we find that he is repeating the assertion of correspondence between poem and building that was originally made by Sir Alfred Clapham on the basis of evidence since shown to be partial and inconclusive.²⁶ Following the authority of Doyle, however, Middle English scholarship on the poem since 1959 has treated the correspondence as fact.²⁷

Interlude: A Ghastly Vignette

When the narrator of *Pierce the Ploughman's Crede* has made the rounds to all four of the mendicant foundations in his imagined streetscape and failed, in each case, to find a friar who could help him learn his creed, he continues his walk outside the city and finds himself in a country scene, face-to-face with working folk whose agricultural labour is in obvious contrast to the idleness of the friars. Like his picture of the mendicants' architectural excess, the poet's portrait of the plowman and his family is, as Clapham would say, graphic:

I seigh a sely man me [by] opon the plow hongen.
 His cote was of a cloute that cary was y-called,
 His hod was full of holes and his heer oute,
 With his knopped schon clouted full thykke;
 His ton todeden out as he the londe treddede,
 His hosen ouerhongen his hokschynes on eueriche a side,
 Al beslombred in fen as he the plow folwede;
 ...
 His wijf walked him with with a longe gode,
 In a cutted cote cutted ful heyghe,
 Wrapped in a wynwe schete to weren hire fro weders,
 Barfote on the bar ijs that the blod folwede.
 And at the londes ende laye a litell crom-bolle,

²⁵ Holsinger, "Lollard Ekphrasis," 84.

²⁶ Doyle, "An Unrecognized Piece of *Piers the Ploughman's Creed*." Doyle's footnote (434 n.35) points to Clapham and Godfrey's *Some Famous Buildings*, quoting the same words that Irwin Smith uses as the basis for the same contention. More recent studies of the London Blackfriars precinct likewise cast doubt on the identification. See for example Rohrkasten, *The Mendicant Houses of Medieval London*, 504; and Holder, *The Medieval Friaries of London*, 43–44.

²⁷ For example, Lawton, "Lollardy in the *Piers Plowman* Tradition," 786, citing Doyle citing Clapham; and Barr, *Piers Plowman Tradition*, 9, 12, and 220 (citing Doyle and Clapham).

And thereon lay a litell childe lapped in cloutes,
 And tweyne of tweie yeres olde opon a-nother syde,
 And alle they songen o songe that sorwe was to heren.

(421–27; 433–40)

Graphic indeed. But would we, would anyone, in reading this passage think for a moment that it describes an actual plowman? That in theory, if one were very assiduous in tracking the evidence, one could *identify* this plowman, if not in name then at least by association with the owner of this particular tillage strip a stone's throw distant from the last of the four mendicant houses the narrator has just visited? Absurd. But to do so would be to follow the same logic that leads us to read the fictional narrator's itinerary as a map, and to read his buildings as real buildings.

When the eminent *Piers Ploughman* scholar and editor George Kane came to this passage in the *Crede* he was characteristically appalled.²⁸ Taking aim at critics who commit the sin of anachronism by bringing modern political sensibilities to bear on fourteenth-century texts which ought to be understood within the tradition of estates satire—and certainly not as examples of what these critics call “dissent”—Kane criticizes the *Crede* poet for getting the details wrong, and his modern readers for taking those details far too literally. Kane paraphrases the portrait of “the wretched peasant” just quoted:

in his coat of course cloth, his hair staring through the holes in his hood, his shoes an imperfect composition of patches, up to his ankles in mud, straining at the plough while his wife, sheltered from the weather only by a winnowing sheet, leaves traces of blood from her bare feet on the ice as she follows the gaunt oxen with the goad, and three hungry children lie whimpering at the end of the tillage strip.

“That ghastly vignette,” he sums up, “is from *Piers the Plowman's Crede*, which is concerned not about politics but about the failure of spirituality.”²⁹ One might not expect, given this insistence on the spiritual over the worldly as the proper frame in which to view the *Crede's* objective, to then find Kane inserting a biting aside about the failure of the poet's realism (in the anachronistic, novelistic sense of the word): “His ragged peasant ankle-deep in mud which, by the way, he has no business trying to plough—*Langland would never have made such a gaffe*—stands for a life in whole accord with spiritual values by contrast with that of self-indulgence exemplified by the friars.”³⁰ This is to say: if you are going to use a touch of the real to seal the deal in an allegorical text, you had better get the details right.

Excursus on Method: Building Blurry

There is a useful analogy to be made, I suggest, between the critical impulse of which Kane disapproves (seeking literal representation, i.e. “description” where that is not

²⁸ Kane, “Some Fourteenth-Century ‘Political’ Poems.”

²⁹ Kane, “Some Fourteenth-Century ‘Political’ Poems,” 85.

³⁰ Kane, “Some Fourteenth-Century ‘Political’ Poems,” 89.

the point), and the impulse in some versions of digital humanities scholarship (and its reception) to create literal representation where that is not the point. I offer my own digital visualization project as an illustration of a tension, as I see it, between the seductive power of technology-driven reconstruction of the past in all its detail, on the one hand, and a deep discipline-driven scepticism and resistance toward the spectre of empiricism that such projects conjure up in the minds of (some) literary critics. Let me be the first to say that I do not pretend to offer an objective position from which to view this debate. I have been seduced by the technology. I am also sceptical of it.

In spring 2010, when Michal Koszycki and I began work on the digital architectural models that would become the basis for our animated film “Great Houses Make Not Men Holy,” I had already completed, several years earlier, an essay outlining the architectural history of the Oxford mendicant foundations in the years between the 1220s (when the friars arrived in England) and the 1530s (when Henry VIII ordered the dissolution of England’s religious houses). Both the film and background essay are available on the Oxford Friars Project website.³¹ It is important to note, though, that while the scholarship gathered in the original essay is typical of the methods used by architectural historians (a synthesis of archaeological evidence with both primary and secondary historical documents), the guiding questions that animated the project were, from the beginning, neither purely historical nor purely architectural. Coming as I did from an English department, my interests in the friars’ buildings were primarily literary. I wanted to know whether the vehement antifraternalism present in late fourteenth-century English writing (most famously in Langland’s *Piers Plowman* and Chaucer’s “Summoner’s Tale,” but extending into Wycliffite prose and poetry, including *Pierce the Ploughman’s Crede*) was reacting to physical changes in the urban streetscapes of English cities, or if these English writers were merely reworking and exaggerating the tropes and conventions of antifraternal critique inherited from Latin precursor texts.³² In the next section I provide a condensed account of the conditions that allowed Chaucer, Langland, and the *Crede*-poet to tap into a longstanding tradition of satirical critique of mendicant building practices—criticism fuelled by tensions within the fraternal orders themselves dating back to days of their founders, Saints Francis and Dominic. This literary history portion of the project, however, could have been accomplished through written argument alone with no digital mediation needed.

Why, then, go to the trouble of making digital architectural models at all? If their purpose is not the painstaking reconstruction of real historical prototypes in virtual space (a worthwhile goal when conditions allow), then what is their purpose?³³

31 Oxford Friars Project, <https://oxfordfriars.wordpress.ncsu.edu/>.

32 The animated film gives brief glimpses of this Latin material in quotations from Matthew Paris (1243) and Richard Fitzralph (1357). See the film script and bibliography at Knowles, Oxford Friars Project. Still the best introduction to the antifraternal tradition is Szittya, *The Antifraternal Tradition in Medieval Literature*.

33 High-profile examples of the “best possible model” approach are Frischer’s Rome Reborn, and the Virtual St. Paul’s Cathedral Project directed by John Wall and colleagues at North Carolina State University.

I suggest that in developing digital models there is equal intellectual value to be gained through taking a “built thought-experiment” approach to the material as compared to taking a “best possible model” approach. For convenience I will call these two approaches “blurry” and “sharp”—though the photographic metaphor is not a perfect one. One argument for building blurry is that reducing the emphasis on verisimilitude as a criterion for the model’s value has the effect of freeing the modeller to emphasize diachronic processes—changes in built environments over time—as opposed to focusing on a static model representing a single moment in time, however perfectly.³⁴ As Caroline Bruzelius has noted: The printed text (the article, the book), as well as the ground plan, section, elevation and photograph, are all representations that capture an isolated moment of production. Print is therefore not a medium conducive to representing the additive process of change over time, or the transformations visited on buildings as a result of the dynamic relationships within cities and/or between institutions ... Animations and reconstructions can model change over time in urban space; they narrate a process as a flow of ongoing interventions, each of which corresponds to the exigencies of a historical moment.³⁵

There are also practical reasons for building blurry versus building sharp. One might not have enough information to render an environment in sharp detail. In the case of the Oxford Greyfriars and Blackfriars models, we had very good archaeological (ground-level) surveys of the sites, and some contemporary accounts confirming their dimensions at various phases of construction, but very little that would count as “description” of the building styles and decorative schemes for either the interiors or the exteriors. We therefore built blurry by necessity, but also out of a sense of intellectual honesty—that including more and sharper architectural detail would have led viewers to think we knew things about these buildings that we did not really know. Other likely practical considerations might be lack of time and lack of money. Building sharp, at scale, is expensive. But lack of access to funding should never be, in my view, a hindrance to building blurry. The questions we can ask and the stories we can tell (and that our students can ask and tell) with free software and a little bit of training, are stories that are worth telling.

Great Houses Make Not Men Holy: The Mendicants in Oxford³⁶

The anonymous Wycliffite tract “Fifty Heresies and Errors of Friars,” dating probably from the early 1380s, begins by pointing out the detrimental effects the friars’ building

³⁴ Compare the discussion by Eileen Gardiner and Ronald G. Musto on the idea of a “construct” (noun): “something that has been created or devised to stand in for or simulate a real-world object, place, process, system or event *over time*.” In Gardiner and Musto, *The Digital Humanities*, 63.

³⁵ Bruzelius, “The Architecture of the Mendicant Orders in the Middle Ages,” 382.

³⁶ This section of the essay is intended as a companion piece to the animated short film “Great Houses Make Not Men Holy: Mendicant Architecture in Medieval Oxford,” produced by Jim Knowles and Michal Koszycki. It is available online as part of the Oxford Friars Project website.

projects have had on parish churches, not just figuratively but in terms of physical damage:

Freris bylden many grete chirchis and costily waste housis, and cloystris as it were castels, and that withoute nede, where-thorw parische chirchis and comyne weyes ben payred, and in many placis undone ... For by this new housinge of freris, thof hit rayne on tho auter of tho parische chirche, tho blynde puple is so disseyved that thei wil rather gif to waste housis of freris then to parische chirchis.³⁷

[Friars build many great churches and costly vast houses, and cloisters like castles, without need, whereby parish churches and common ways are impaired and in many places undone ... Because of this new housing of friars, even though it rains on the altar of the parish church, the blind people are so deceived that they would rather give to vast houses of friars than to parish churches.]

The charge here is that donations to the friars have diverted funds from the upkeep of parish churches to such an extent that it is raining on the altars. William Langland employs the same striking image in the C version of *Piers Plowman*, also from the 1380s, where he laments the fate of church properties formerly attached to private estates which have been donated to the religious orders:

Lytel hadde lordes a-do to yeve lond fro her heyres
To religious that haen no reuthe thow it ryne on here auters.

[Lords had no business giving land from their heirs to religious, who have no compunction even though it rains on their altars.]³⁸

Here the critique is not limited to the friars (monks are guilty too), but taken together these two quotations are indicative of just how far popular conceptions of the mendicant orders had shifted in the short century and a half since their arrival in England—since “Franceys tyme” [Francis’s time], when, according to Langland, Charity itself had worn a friar’s frock.³⁹

Given the situation in late fourteenth-century England, there is considerable historical irony in the fact that Saint Francis’s earliest vision of the new apostolic life came in the form of an architectural mandate from the crucified Christ: “Francis, go rebuild my house.” Bonaventure, who tells the story in his *Minor Legend* (1263), characteristically interprets Francis’s early building activity allegorically as the repairing of the mystical

³⁷ The full tract is reprinted in Arnold, *Select English Works of John Wyclif*, vol. 3, 366–401; quotation at 380. For the likelihood that this piece and a high proportion of early Wycliffite material in the vernacular originates in Oxford, see Hudson, *The Premature Reformation*, 82–110.

³⁸ Langland, *Piers Plowman*, passus 5, lines 163–64.

³⁹ Langland, *Piers Plowman*, passus 16, lines 352–53.

body of Christ. The founder's own first reaction to the mandate was more materialistic; he immediately began to mend the dilapidated church of San Damiano in which he had received the vision.⁴⁰ Long before Bonaventure's redaction of the legend, however, several of the founding documents of the order had shown a similar preoccupation with material spaces (or lack thereof) and the new brothers' imagined relations to them. The Later Rule (*Regula Bullata*) of 1223, for example, forbids the brothers to acquire possessions of any kind, but singles out only "houses" and "places" for special emphasis: "The brothers shall not acquire anything as their own, neither a house nor a place nor anything at all. Instead, as pilgrims and strangers in this world who serve the Lord in poverty and humility, let them go begging for alms with full trust."⁴¹ Francis's final *Testament* of 1226 picks up the "as pilgrims and strangers" simile—a reference to 1 Peter 2:11—but replaces the prohibition on "houses" and "places" with the more specific "churches or poor dwellings." Yet this apparent move toward a more restrictive interpretation of the Rule is immediately tempered by a qualifying clause which moves in an opposite and, from the perspective of architectural history, an enabling direction. "[U]nless," the clause reads, "[the dwelling] is in harmony with [that] holy poverty which we have promised in the Rule"—thus allowing for historically dynamic and shifting interpretations of the "holy poverty" that Francis saw as a precondition for any building project whatsoever.

For their part, the Dominicans had a less problematic relationship to building and to property in general than did the Franciscans, but here too the early documents are characterized by conflicting attitudes within the order.⁴² The *Acts of Canonization of Saint Dominic* report the founder's distress upon learning of improved living conditions for the friars at Bologna: "Do you wish to give up poverty so quickly," he asks, "and build great palaces?"⁴³ Ornamentation was also subject to strictures. In 1245 an act of the general chapter of the order attempted to limit the scale of "sepulchral monuments" within the Preachers' churches, and in 1250 another act tried to proscribe such burials altogether.⁴⁴ That such acts failed to achieve their desired result is beyond debate, but a number of questions remain as to how and why the mendicant orders moved from a self-conscious and self-limiting posture to one of purported architectural excess. However distinct their original motives and methods may have been, by the late fourteenth century the Franciscans and Dominicans found themselves lumped together in the kinds of architecturally targeted critiques with which I began this section, and of which *Pierce the Ploughman's Crede* may have been, as Clapham claimed, the most graphic example. How had this happened? What was the nature of these "grete chirchis" and how had their builders reconciled their construction and maintenance with the mendicant ideal of apostolic poverty? And were these churches actually as "grete" in terms of size and opulence as their detractors claimed? In

40 Bonaventure, *Minor Legend*, 684–717.

41 Francis of Assisi, in Armstrong and Brady, ed., *Francis and Clare*, 141.

42 Hinnebusch, *The Early English Friars Preachers*, 121–30.

43 Hinnebusch, *The Early English Friars Preachers*, 126.

44 Hinnebusch, *The Early English Friars Preachers*, 127.

answering these questions specifically with regard to these two orders' sites in Oxford, our digital models, however blurry, succeed in showing that the Oxford houses of the Franciscans and Dominicans contributed disproportionately to the overall negative picture painted of the mendicants by English writers in the later fourteenth century. The still images in the paragraphs to follow are taken from our animated film and from the digital models underlying it. As such, they provide a limited case study of what can be accomplished using inexpensive modelling software (SketchUp in this case) and a blurry building approach to rendering historical evidence in the service of a larger narrative. And although it cannot be reproduced in print, the additional layer of digital mediation provided by the animation sequence is crucial to the overall effect of the project. As Bruzelius argues, animations are of special importance to mendicant architectural history, because it was "almost by definition about change over time. Not only are animations an effective way to demonstrate progressive transformation as a function of time and space, but they provide a means to engage both with building as ongoing process and to model evolving and dynamic relationships between religious institutions."⁴⁵

The first structure on the Franciscan site for which we have material remains is the second chapel.⁴⁶ The first recorded chapel, which was begun in 1232 and demolished sometime before 1272, must have stood inside the wall and to the south of St. Ebbe's church, but excavations made in this area between 1967 and 1976 found no traces of this structure.⁴⁷

Construction on the second chapel began in 1244 and was probably finished by 1248 (see Figure 4.1). This was a simple rectangular structure measuring approximately 100 by 30 feet, with no structural division between the choir and nave. Of extant Franciscan churches this building is most comparable to the upper church at Lincoln, which was built at around the same time.⁴⁸

Phase II in the reconstructed plan comprises a major expansion of the building to the north and west, the original chapel being retained as the eastern portion of the building

⁴⁵ Bruzelius, "Architecture of the Mendicant Orders," 382.

⁴⁶ For the basic outline of the Franciscans' arrival and early history in Oxford, I have relied on Little, *The Grey Friars in Oxford* and his article "The House of Greyfriars" in the *Victoria County Histories*. Despite their age, these pieces have not been superseded as synthetic accounts of these events. Details are supplemented by the account in Hassall, Halpin, and Mellor, "Excavations in St. Ebbe's, Oxford, 1967–1976." The Greyfriars excavations conducted by Hassall et al. from 1967–76 are now to be superseded by new and more extensive excavations carried out by Oxford Archaeology in 2015. Details are available at: <https://oxfordarchaeology.com/our-projects/westgate-oxford>.

⁴⁷ Hassall et. al, "Excavations in St. Ebbe's," 184. Almost nothing remains above ground for any of the building phases discussed below. According to Hassall, "the only surviving pier bases uncovered in the [Greyfriars] excavations have been retained in a shop basement" (email correspondence December 2014). My discussion is therefore dependent on the reconstructed (and incomplete) priory plan created by Hassall et al. (1989), which is based entirely on foundation-level evidence. The 3-D models featured in the animated film are based on Hassall's reconstructed ground plan.

⁴⁸ Hassall et. al, "Excavations in St. Ebbe's," 186. See discussion and photos in Martin, *Franciscan Architecture in England*, 89–100.

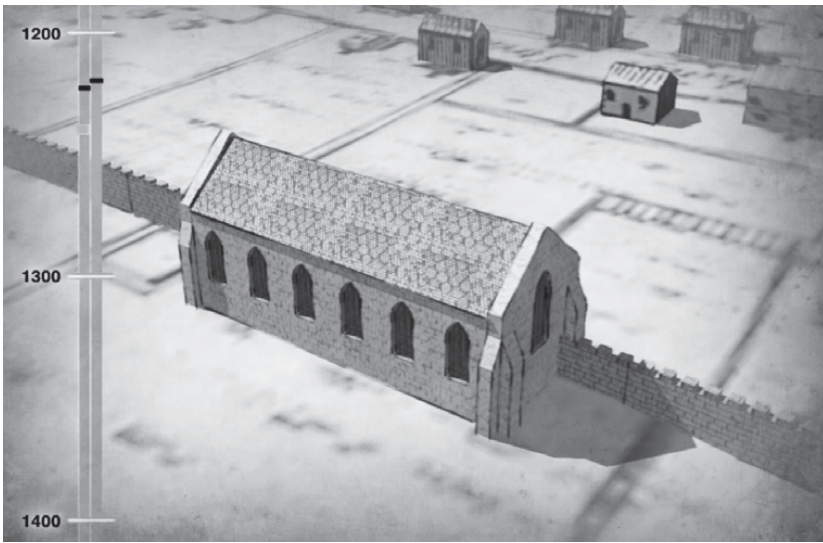


Figure 4.1. Greyfriars second chapel, Phase I, 1244–48.

and now designated as the friars' choir (see Figure 4.2). The new western nave included an arcaded north aisle approximately 10 feet wide, making the width of the church 40 feet at its west end. At its east end the single north aisle abutted another new addition at this phase, a small north "transept" (so named to distinguish it from the later north "nave"). Based on surviving pier bases and mouldings, T. G. Hassall's 1989 archaeological survey confidently dates these additions to before 1270. However, the exact relationship between the termination at the east end of the new northern aisle and the new north transept remains uncertain; there was no evidence discovered in the ground for the postulated wall separating the north aisle from the north transept, or the extension of this wall separating the crossing from the nave.⁴⁹

The most extensive set of additions to the Greyfriars church was yet to come. The years between 1270 and 1330 encompassed two major building projects, here labelled phases III and IV (Figures 4.3 and 4.4). Dates are uncertain for both phases. The Franciscans first extended the western end of the nave, including the north aisle, by approximately 30 feet, making the length of the nave 110 feet.⁵⁰ Moreover, the new walls at the northwest corner of the church appear heavily reinforced, as if designed to bear more than just the weight of the roof. Two possibilities present themselves, neither with supporting documentary evidence. First, it seems likely that this area of the church served as the main entrance for the laity, and thus the reinforced walls may have served to support a new façade or porch added at this stage. The removal of the bell tower from the centre of the church

⁴⁹ Hassall et. al, "Excavations in St. Ebbe's," 186.

⁵⁰ Hassall et. al, "Excavations in St. Ebbe's," 188.

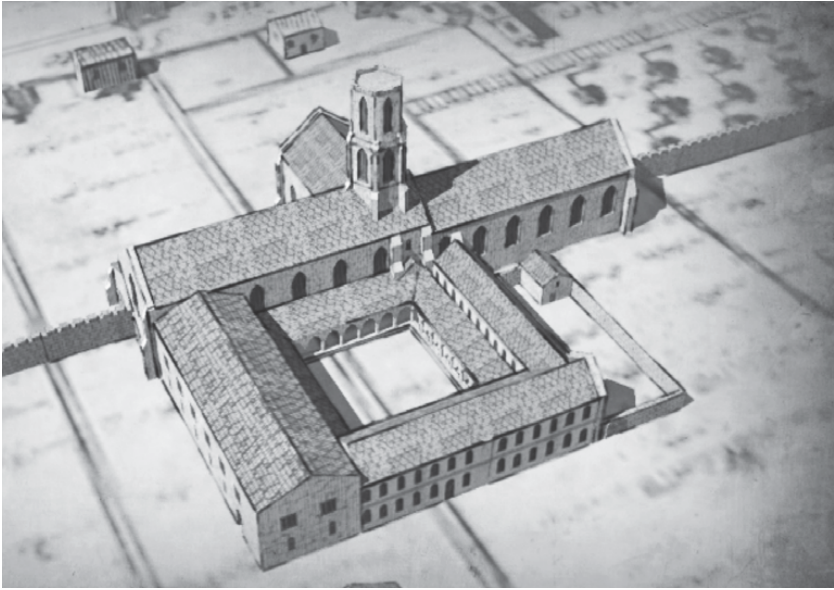


Figure 4.2. Greyfriars, Phase II, 1260–70.



Figure 4.3. Greyfriars, Phase III. Blackfriars is visible in the background.

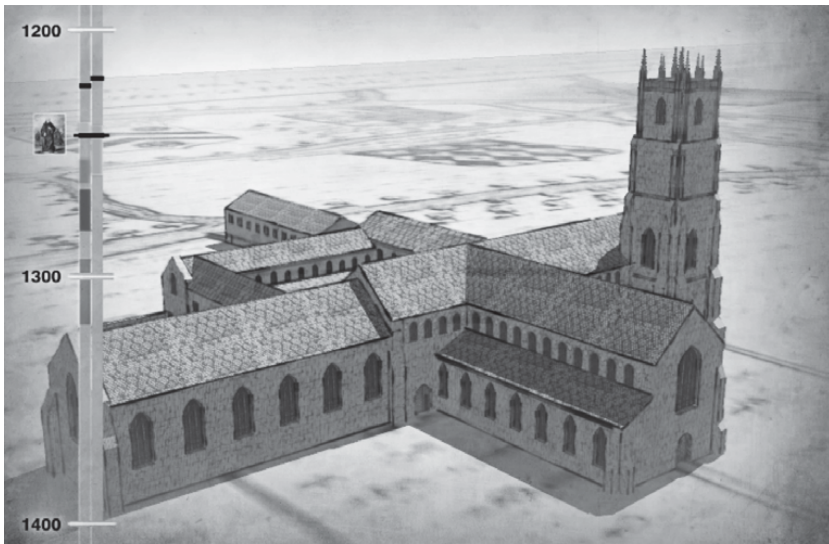


Figure 4.4. Greyfriars, Phase IV, after 1270.

would have allowed for a more spacious and open floor plan following completion of the northern nave.

Phase IV involved much more than a simple extension of the existing transept. Rather, the western wall of the new nave was shifted two bays to the west, while the eastern wall was shifted one bay to the east, resulting in wider vessel that incorporated none of the existing transept walls. The new nave was then extended north to more than double the length of the old transept, and seven private chapels were added along its eastern wall.⁵¹

The next phases of building are difficult to date with precision. Sometime between 1330 and the beginning of the Black Death (1348–49), the north nave was extended by an additional three chapels, making a total of ten. Around the same time, part of the north wall of the north aisle was rebuilt. This construction may be associated with a gift of stone from Edward III in 1346.⁵² Almost nothing is known about the contents or patrons of the ten chapels. Little suggests that “they would be in part sepulchral chantries, supported by noble families or guilds, often containing the image or shrine of some saint.” “In the thirteenth century,” he continues, touching on a familiar theme, “benefactors, great and small, were willing and eager to help the friars to raise those splendid buildings which drew forth the fierce denunciations of later reformers.”⁵³

⁵¹ Hassall et. al, “Excavations in St. Ebbe’s.”

⁵² Hassall et. al, “Excavations in St. Ebbe’s”; Little, “The House of Greyfriars,” 125.

⁵³ Little, *The Grey Friars in Oxford*, 24–25



Figure 4.5. Greyfriars, Phase V, late fourteenth century.

A reconstructed plan of the church at its largest, around 1350, can be seen in [Figure 4.5](#). Archaeological evidence shows what appears to be some retrenchment dated after 1480, when the north nave was walled off from the older nave and choir and apparently allowed to fall into ruin.⁵⁴ No historical explanation for this retrenchment is yet forthcoming, but the irony of a roofless Franciscan nave, its altars exposed to the rain, would not have been lost on Langland and the *Crede* poet of a century earlier.

In contrast to the Franciscans, who expanded their land holdings and their buildings in a number of phases over a hundred-year period, the Dominicans in Oxford planned and executed the large church and convent on their second site in one concentrated burst of activity between 1236 and 1245 (see [Figure 4.6](#)).

Archaeological work conducted on the second Blackfriars site between 1961 and 1985 has revealed a very ambitious building plan for the new church and priory.⁵⁵ As the reconstructed plan makes clear, the church, encompassing a double-aisled nave and central walking-place, the friars' choir to the east, and the main cloister precinct (including a very large chapter house) were all constructed within a short span of time in the middle of the thirteenth century. (Note that the overall precinct layout is nearly identical to the plans proposed for London Blackfriars by first Clapham, then Adams and Smith—plans they were convinced matched the description in *Pierce the Ploughman's Crede*). The cloisters were doubled sometime probably in the mid-fourteenth century. The only

⁵⁴ Hassall et. al., "Excavations in St. Ebbe's," 191.

⁵⁵ Lambrick and Woods, "Excavations on the Second Site of the Dominican Priory, Oxford," and "Further Excavations on the Second Site."

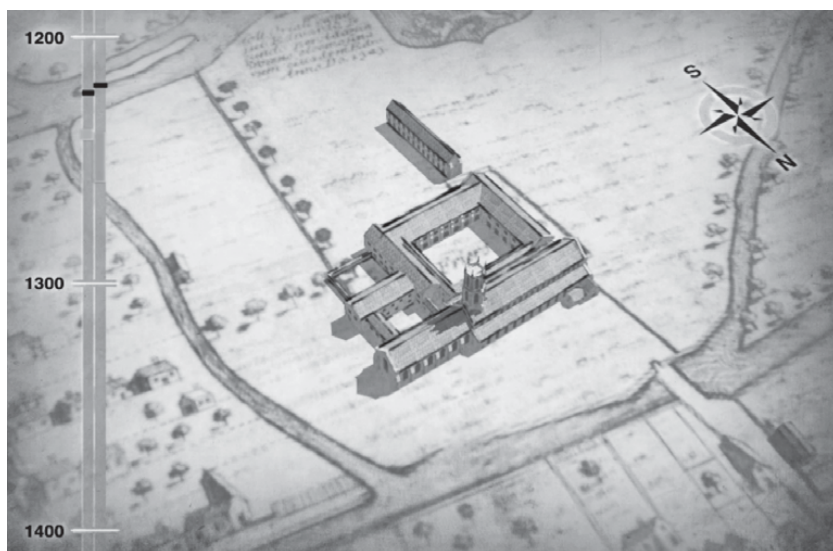


Figure 4.6. Blackfriars second site, Phase I, 1236–45.

identifiable later additions to this basic plan were two extensions to the western end of the nave, dating probably from the early fifteenth century (see Figure 4.7).⁵⁶ The completed church measured 240 ft. in length, making it the largest known Dominican church in England (10 ft. longer than London Blackfriars, completed 1286).

To conclude this section, I wish to draw attention to the larger perspective shown in Figure 4.8, where the pre-Reformation boundaries of the Franciscan and Dominican precincts are outlined on Ralph Agas's 1578 map of Oxford. In the later Middle Ages, Oxford was still a relatively small college town. It is striking to view in this perspective the sheer enormity of the friars' physical presence in the southwest part of the city following the completion of their building projects. The size of their churches and cloisters, as the models show, taken along with the combined size of their walled precincts in comparison to the city as a whole, make it immediately apparent why the friars came to be treated with a combination of awe, admiration, and distrust.

In conjunction with the history of scholarly speculation attached to the "graphic architectural description" in *Pierce the Ploughman's Crede*, outlined above, I hope it is also apparent that the scale and opulence of the Oxford priories poses a serious challenge to the position that London was the only reasonable location in which to situate a satirical critique of mendicant architectural excess in the late fourteenth century. It would go against the grain of my story, though, simply to replace London with Oxford as the proper site of the *Crede* narrator's itinerary. I wish to make no such claim. I want only to suggest, borrowing a favourite phrase of the Dominican theological school at

⁵⁶ Lambrick and Woods, "Excavations on the Second Site of the Dominican Priory, Oxford," 189–90.



Figure 4.7. Blackfriars, Phase II (foreground). Greyfriars, Phase V (background).



Figure 4.8. The underlying image is a digital copy of the Ralph Agas 1578 map of Oxford, a copy of which is in the Bodleian Museum. For a digital copy see: <https://digital.bodleian.ox.ac.uk/inquire/p/366cb8e3-360d-4252-b41e-7a55fd806cf>.

Oxford (whose members lived and worked in these buildings), that “it could have been otherwise.”⁵⁷

Conclusion: A Ghast in the Machine

The digital reconstruction project outlined briefly in the previous section was built on evidence provided by historians and archaeologists whose work undergirds the story we set out to tell. That I can call that work “evidence,” while at the same time criticizing, in “The Ghost of Shakespeare’s Blackfriars,” above, the use of the same word by scholars who wish to appropriate the ekphrastic poetry of *Pierce the Ploughman’s Crede* to suit the interests of the Shakespeare industry, seems to me symptomatic of the dilemma posed by bringing new technologies to bear on interpretations of the past. Is our film a digital version of George Kane’s “ghastly vignette”? The answer is yes, in two senses. It is self-evidently ghastly (Middle English for ghostly) in that it is an immaterial artefact, constructed out of thin air, or rather bits and bytes. And it is a vignette because, like Kane’s plowman, it is a stand-in for the real, a masquerade, a creative, analogical, and allegorical response to the material past, not a verifiable reconstruction of it. The danger here, says the sceptic, is that the reified visible model on the computer screen will be mistaken for the primary source, for the thing itself, when it actually is another layer of interpretive filter that colours our picture of the past but does not in any way get us closer to it. I have been repeatedly surprised when presenting the film to audiences (including to professional medievalists) by how often the question of empirical verisimilitude comes up. “Can’t you go to Oxford with a tape measure,” I have been asked, “and figure out *exactly* how wide that archway really was?” Maybe I could, but to do so would be to commit a gaffe even worse than the one Kane accuses the *Crede* poet of committing (who no doubt knew that ploughing in the mud was impossible, but couldn’t resist the sheer ghastliness of it). I would prefer to commit the gaffe of not getting it right—of not even pretending that I could get it right. Like the *Crede*’s ploughman, our digital ghosts are not intended to reconstruct the past, but to ask questions of it, and to be questioned by it. I will close where I began, by quoting Lee Patterson, who did not have the digital humanities in mind when he wrote “Historical Criticism and the Claims of Humanism,” but it sounds as though he might have seen it coming. It sounds too like he would have agreed with me about the virtues of building blurry: “Rather than conceive of history as a completed product capable of being examined by a detached observer, the alternative is instead to recognize that the unfinished nature of the historical process, however disabling to the quest for methodological correctness, endows the critic’s activity with historical consequence: the past we reconstruct will shape the future we must live.”⁵⁸

⁵⁷ See Gelber, *It Could Have Been Otherwise*.

⁵⁸ Patterson, “Historical Criticism,” 45.

Bibliography

- Adams, Joseph Quincy. "The Conventual Buildings of Blackfriars, London, and the Playhouses Constructed Therein." *Studies in Philology* 14 (1917): 64–87. www.jstor.org/stable/4171701.
- Arnold, Thomas, ed. *Select English Works of John Wyclif*. 3 vols. Oxford: Clarendon Press, 1871.
- Barr, Helen, ed. *The Piers Plowman Tradition: A Critical Edition of "Pierce the Ploughman's Crede," "Richard the Redeless," "Mum and the Sothsegger," and "The Crowned King."* London: Dent, Everyman's Library, 1993.
- Bonaventure, Saint. *The Minor Legend of Saint Francis*. In *Francis of Assisi: Early Documents*. 3 vols. Edited and translated by Regis Armstrong, J. Wayne Hellmann, and William J. Short. New York: New City Press, 2000.
- Bruzelius, Caroline. "The Architecture of the Mendicant Orders in the Middle Ages: An Overview of Recent Literature." *Perspective* 2 (2012): 365–86.
- Clapham, Alfred W. "On the Topography of the Dominican Priory of London." *Archaeologia* 83 (1912).
- Clapham, Alfred W., and Walter H. Godfrey. *Some Famous Buildings and Their Story; Being the Results of Recent Research in London and Elsewhere*. London: Westminster Technical Journals, 1913.
- Doyle, A. I. "An Unrecognized Piece of *Piers the Ploughman's Creed* and Other Work by Its Scribe." *Speculum* 34 (1959): 428–36.
- Francis of Assisi, Clare of Assisi. *Francis and Clare: The Complete Works*. Edited and translated by Regis Armstrong and Ignatius Brady. New York: Paulist Press, 1982.
- Frischer, Bernard. Rome Reborn. <http://romereborn.frischerconsulting.com>.
- Gardiner, Eileen, and Ronald Musto. *The Digital Humanities: A Primer for Students and Scholars*. Cambridge: Cambridge University Press, 2015.
- Gelber, Hester. *It Could Have Been Otherwise: Contingency and Necessity in Dominican Theology at Oxford, 1300–1350*. Leiden: Brill, 2004.
- Gurr, Andrew. "London's Blackfriars Playhouse and the Chamberlain's Men." In *Inside Shakespeare: Essays on the Blackfriars Stage*, edited by Paul Menzer, 17–30. Selinsgrove: Susquehanna University Press, 2006.
- Hassall, T. G., C. E. Halpin, and M. Mellor. "Excavations in St. Ebbe's, Oxford, 1967–1976: Part I: Late Saxon and Medieval Domestic Occupation and Tenements, and the Medieval Greyfriars." *Oxoniensia* 54 (1989): 72–277.
- Hinnebusch, William, O. P., *The Early English Friars Preachers*. Rome: Istituto Storico Domenicano, 1951.
- Holder, Nick. "The Medieval Friaries of London: A Topographic and Archaeological History, Before and After the Dissolution." PhD dissertation, University of London, 2011.
- Holsinger, Bruce. "Lollard Ekphrasis: Situated Aesthetics and Literary History." *Journal of Medieval and Early Modern Studies* 35 (2005): 67–89.
- Hudson, Anne. *The Premature Reformation: Wycliffite Texts and Lollard History*. Oxford: Oxford University Press, 1988.

- Kane, George. "Some Fourteenth-Century 'Political' Poems." In *Medieval English Religious and Ethical Literature: Essays in Honour of G. H. Russell*, edited by Gregory Kratzmann and James Simpson, 82–91. Cambridge: Brewer, 1986.
- Knowles, Jim. The Oxford Friars Project. <https://oxfordfriars.wordpress.ncsu.edu>.
- Lambrick, George, and Humphrey Woods. "Excavations on the Second Site of the Dominican Priory, Oxford." *Oxoniensia* 41 (1976): 168–231.
- . "Further Excavations on the Second Site of the Dominican Priory, Oxford." *Oxoniensia* 50 (1985): 131–208.
- Langland, William. *Piers Plowman: A New Annotated Edition of the C-Text*. Edited by Derek Pearsall. Exeter: University of Exeter Press, 2008.
- Lawton, David. "Lollardy in the *Piers Plowman* Tradition." *Modern Language Review* 76 (1981): 780–93.
- Little, A. G. *The Grey Friars in Oxford*. Oxford: Oxford Historical Society, 1892.
- . "The House of Greyfriars." In *Victoria County Histories, Oxon.*, 122–37. London: Archibald Constable, 1907.
- Martin, A. R. *Franciscan Architecture in England*. Manchester: Manchester University Press, 1937. Reprint 1966 by Gregg Press.
- Patterson Lee. "Historical Criticism and the Claims of Humanism." In *Negotiating the Past: The Historical Understanding of Medieval Literature*. Milwaukee: University of Wisconsin Press, 1987.
- Rohrkasten, Jens. *The Mendicant Houses of Medieval London, 1221–1539*. Münster: LIT, 2004.
- Simpson, James. *Reform and Cultural Revolution*. Vol. 2 of *Oxford English Literary History, 1350–1547*. Oxford: Oxford University Press, 2002.
- Skeat, Walter W., ed. *Pierce the Ploughmans crede (about 1394 A.D.) transcribed and ed. from ms. Trin. coll., Camb., R. 3. 15*. Early English Text Society, o.s., 30. Oxford 1867. Reprint 1895.
- Smith, Irwin. *Shakespeare's Blackfriars Playhouse: Its History and Design*. New York: New York University Press, 1964.
- Szittyá, Penn. *The Antifraternal Tradition in Medieval Literature*. Princeton: Princeton University Press, 1986.
- Todd, James Henthorn, ed. *An Apology for Lollard Doctrines attributed to Wycliffe*. London: Camden Society, 1842.
- Wall, John, et al. Virtual St. Paul's Cathedral Project. <https://vpcp.chass.ncsu.edu>.
- Warton, Thomas. *A History of English Poetry: From the Close of the Eleventh Century to the Commencement of the Eighteenth Century*. 3 vols. London: Printed for J. Dodsley, 1774–81.

Chapter 5

CONTENT IS NOT CONTEXT: RADICAL TRANSPARENCY AND THE ACKNOWLEDGEMENT OF INFORMATIONAL PALIMPSESTS IN ONLINE DISPLAY

MATTHEW EVAN DAVIS

WHETHER FOR PRINT or for the internet, dividing structure and content—the layer approach used in modern web development—has influenced our modern notions of textual presentation. Conscious of it or not, popular conceptions of “content” treat the text as a Platonic ideal floating in the cloud, divorced from any mechanisms of production or display. Since the presentation and display layers are handled separately in most modern web and publishing tools, the underlying assumption is that content can fluidly fit any container it is placed into, like water poured into beakers of differing shape, but similar volume. As scholars of medieval manuscript and early print culture can attest, however, this is ultimately a dangerous misconception. For example, in this very volume Timothy Stinson has pointed out that the act of “translating” medieval scribal texts to printed works has “profoundly shaped conceptions of medieval authorship and textuality and coloured the way we understand, read, and teach medieval literature.”¹ How much more, then, does the separation of presentation and display alter our understanding? Likewise, Tamsyn Mahoney-Steel’s chapter notes that, even when a single manuscript exists “the loss of information in the translation from parchment to page or screen, is still great.”² If the philosophy behind the modern notion of “content” is true—that it fluidly fits whatever space we wish it to in whatever manner we want—then surely the medieval manuscript and its print editions should be able to do so as well. As Mahoney-Steel’s cogent statement on the loss of information points out, however, this is not the case.

The reality is that any action taken to inscribe text—whether the initial act of creation, an act of interpretation, or an act of presentation in a manuscript, printed book, or on an online display—is inherently an act of editorial interpretation at best and intervention at worst. The tools, infrastructures, and methods we use—and, increasingly, the standards we attempt to enfold all texts within under the banner of interoperability—have certain expectations and goals in mind, often built around the metadata ontologies used to allow text to be read by a machine and the needs of the software development cycle. Those goals may or may not correspond to the researcher’s goals in developing a virtual archive or those of the original authors, scribes, and editors of the manuscripts the tool is working with. Instead, these tools and methods are largely a black box, defined here as anything that receives input and generates output but does not allow the observer to discern its underlying workings.

1 Stinson, “Of Dinosaurs and Dwarves.”

2 Mahoney-Steel, “Encoding and Decoding Machaut.”

The existence of such black boxes in scholarship, especially digital scholarship, are known and worked against.³ However, as more scholars are encouraged by institutional direction and funding opportunities to pursue the development of tools and methods for the digitization and display of medieval manuscripts it is worthwhile to state starkly that the development of a tool or methodology is every bit as much of a theoretical action as the use of theory in writing.⁴ Moreover, the more we digitize and make available medieval texts online the more likely it is that undergraduates, masters and doctoral students, and our colleagues from financially strapped institutions will first encounter manuscripts through these virtual facsimiles—with consequences for their scholarship and understanding of the material object that the virtual facsimile adapts for display on the web. Thus, the choices we make now—often imposed upon us by programming infrastructures, prior example, institutional fiat, or the exigencies of the funding cycle—have very real consequences that may become baked into our thinking.

Beginning with a brief description of the history of the commonly understood file/folder computing metaphor, this chapter will describe the author's thinking about how to display the digital object online. It draws largely from his own consideration of the intersection between Actor-Network Theory and Peircean Semiotics, taken broadly from Carl Knappett's work in material culture and archaeology and interpreted alongside the necessities of religious performance and adaptation as a mediating network of ideas and methods in which texts are read, interpreted, and re-inscribed.⁵ Because readers tend to ignore this background of ideas and methods, the network appears as a unitary thing that a reader treats as a single "text": a process called punctualization. However, when errors arise this illusion breaks down and the tensions between the various aspects of the network that made up that punctualized "text" are often unproductively and haphazardly revealed. If we purposely create channels by which a reader or viewer naturally experiences the collapse of this mediating network and the tensions it produces—a process I call "radical transparency"—both the digital or virtual facsimile as an adaptation of the analogue original and that analogue original itself are no longer treated as exactly the same thing. Instead they are connected, but unique artefacts, and the paratexts of those artefacts are therefore foregrounded.

The chapter will then situate the mediating network and possible avenues of intended collapse and radical transparency within the bounds of methodologies and tools for digital presentation of text and image online. It does so chiefly through describing the practical aspects of developing the author's primary digital project—the Minor Works of John Lydgate Virtual Archive. It is thus intended to articulate some of the theoretical

3 See, for example, Grant Glass's blog post on the black box, Gibbs and Owens "Building Better Digital Humanities Tools," and Nowvieskie's call for the teaching of algorithmic thinking in the humanities "A Game Nonetheless."

4 Ramsay and Rockwell, "Developing Things."

5 Knappett, "Networks of Meaning." While my own interpretation of the relationship between material culture, religious performance, and adaptation largely begins with this chapter, the preceding four would also be worth reading as an introduction to concepts that Knappett discusses in depth here.

underpinnings of the author's work with fifteenth-century manuscripts of John Lydgate, but more importantly to speak to a larger concern with praxis in the development of digital humanities tools and the necessity to underscore what is an oft-ignored truism of digital humanities work: that tools and methodologies, whether developed by a single author or collaboratively, are themselves theoretical actions that are incorporated into the mediating network. They must be explained and justified transparently, as one would any theoretical approach used in scholarship. To make the assumption that the tools used to build platforms for display and comparison are themselves theoretically agnostic abstracts the primary mechanism of mediation for the medieval manuscript on the web when that does not need to be the case. That abstraction, in turn, can cause decisions made on the spur of the moment or because of technical limitations of the tools and methods available to become entrenched modes of thinking, and from those entrenched modes ultimately to distort the primary or only interpretive model accessible to most scholars.

In articulating my concern about the ways in which our tools embody ideologies that, in turn, shape culture in unexpected ways I am certainly not alone. Langdon Winner, Peter Kropotkin, William Morris, and Lewis Mumford all wrote pieces either critiquing the ways that mass industrialization—the nineteenth-century version of interoperability—was shaping culture detrimentally or the ways in which architecture, urban planning, and the development of physical, analogue objects embed aspects of the cultural ethos of their producers for both good and ill. Environmental activists, in turn, often used these theories in the debates between the use of nuclear and solar power in the 1970s and 1980s, with the latter framed as the “soft,” democratic option and the former as the tool of a totalitarian state.⁶

Speaking more specifically to the ways in which the digital shapes our culture, Lisa Nakamura has written extensively on the democratizing nature of the internet and the ways in which it allows a separation between body and self, a separation that she speaks of utopianly in her earlier works but that seems to have never truly been realized with the ascendancy of social media.⁷ Likewise, Wendy Hui Kyong Chun has noted in her recent *Updating to Remain the Same* that much of the work of digital culture is really the result of habit. She states that “through habits users become their machines: they stream, update, capture, upload, share, grind, link, verify, map, save, trash, and troll. Repetition breeds expertise, even as it breeds boredom.”⁸ Moreover, in describing the connection between repetition and habit, she points out that “habit supports a worldview driven by automation and automatic codes, which reduces the future to the past, or more precisely, a past anticipating the future.”⁹ This repetition of habit—the process of

⁶ Winner, “Do Artifacts Have Politics,” 121.

⁷ Nakamura, *Digitizing Race*.

⁸ Chun, *Updating to Remain the Same*, 1.

⁹ Chun, *Updating*, 70. Chun's idea of habit has connections with the idea of ritual—although there without the pejorative of “boredom”—as a marker of culture, as well, and her invocation of code as a reinforcement of habit invokes the ways that ritual can become encoded as memes to be

constant inscription and re-inscription—is what she considers a hallmark of digital culture. She describes it as flat, with the past, future, and now reduced to a single moment of “upgrade,” and the lack of history as anything beyond the eternal now means that the objects of our study are in constant danger of being swept up into a black box of simply being “content”—grist for the cycle of update, process, repeat that Chun speaks of.

What makes this danger starker is that while the nineteenth-century critics and even the environmental activists of the 1970s and 1980s wrote for a public or semi-public audience, critics like Chun or Nakamura are often writing—intentionally or not—for an audience of their fellow academics in either their own or adjacent and allied disciplines. The audience that is developing these tools, and thus shaping the culture, are not thinking about the ways that their tools embody habit or crisis as Chun discusses them—except, perhaps, in how habit and crisis can be easily monetized, as occurs in social media. They are certainly not thinking about the implications: as Sean Parker noted in a November, 2017 Axios.com event regarding the creation of Facebook “I don’t know if I really understood the consequences ... of a network when it grows to a billion or 2 billion people ... it literally changes your relationship with society, with each other.”¹⁰

Parker’s quote is telling because the boundaries that separate both formal academic software engineering and, more importantly, the code mills and bootcamps replacing formal academic training as the means by which programmers understand digital tools and their role in their creation and promulgation, from the humanities are stark. Assuming that both halves of the partnership to develop new tools, platforms, and methods have the same concerns in developing the tool is naïve at best and dangerous at worst. Parker has gone on to become “something of a conscientious objector” on social media platforms like Facebook,¹¹ but the tool—and the ideas embodied in the development of the tool—have become part of our overarching culture. If, as humanists, we lack the public audience of the environmental activists of the 1970s and 1980s or the nineteenth-century critics of industrialization then our understanding of the material object must be encoded directly into the tools we use. Otherwise, the archival equivalent of Facebook—and the flat version of history and the object that it promotes—will do it for us.

User Design as Theory: the File/Folder Metaphor, the Entrenchment of Ideas, and Unintended Consequences

Because of their ubiquity in our daily lives, the technologies we work with are not often considered a theoretical apparatus, yet that is exactly what they are. Take, for example, the concept of a computer file. The mental shorthand referring to it as such first appears in a 1950 RCA vacuum tube advertisement, with the title “Tube with a memory keeps

repeated according to specific, set standards that must be referred back to and compared against for correctness. See Drout, *How Tradition Works* for further information.

¹⁰ Allen, “Sean Parker Unloads on Facebook.”

¹¹ Allen, “Sean Parker Unloads on Facebook.”

answers on file.”¹² This made a degree of sense at the time as the physical hardware, rather than the information, was being referred to. The word “file” was used because the mechanism by which that information was kept was functionally akin to the familiar physical files that kept information in analogue filing cabinets.¹³ Thus, when the first iterations of the UNIX operating system were created in the early 1970s the computer programs and sets of information created for it were also called files. As were the computer systems and sets of information for the first personal computers to find mass usage in the home market in the late 1970s and early 1980s. This concept—the “file” as metaphor for a packet of connected information—has continued to the present day as an easily understood and arguably no longer necessary shorthand.¹⁴

Thinking beyond the concept of “file” and “folder” as shorthand, however, the reference to information stored on a computer as “files” has not just affected information storage, but also the development and design of operating systems as the means through which a user imposes their intent upon the machine. Most modern operating systems store information in a series of nested “folders,” in which each “file” sits. Although sometimes these larger sets are called “directories,” they are in fact functionally similar to a large file folder containing many smaller files and folders, arranged in drawers (directories), and accessed through a system that is similar to the finding systems of any large library or other analogue information repository.

There is no reason why information has to be stored in this manner anymore, and the idea of “files” is a mental conceit to allow users who were steeped in analogue file systems to get up to speed on the computer quickly. In fact, the earliest forms of the pre-OS X Macintosh File System were “flat”—that is to say, they had all of their information stored at what we would consider the root of the filesystem.¹⁵ Even this version of the operating system was a compromise. As Jaron Lanier notes:

The first iteration of the Macintosh, which never shipped, didn’t have files. Instead, the whole of a user’s productivity accumulated in one big structure,

12 Radio Corporation of America, Advertisement, 96.

13 In fact, the overarching standard for UNIX-based computer systems is called the Filesystem Hierarchy Standard—a reference to a piece of hardware intended to organize physical “disk files”: in essence, disk drives—as well as multiple virtual files on a single storage device in the early 1960s.

14 This use of metaphor to discuss the ways in which information is encoded, decoded, and understood is not unique to the digital age, obviously. In *The Book of Memory*, for example, Mary Carruthers notes that the concept of memory for much of history was analogous to the process of inscription, and that the recall of those images occurred in the same way that notes were often recorded onto wax tablets, to be retrieved later—a process that is very similar to the metaphor of computer data as files and folders that RCA invoked in their ad copy (17). She also goes on to point out that additional information—the network surrounding the actual item to be recalled—could be used as a means of situating memory in a way that evokes our own attempts to create a semantic web (61).

15 This concept has been revived by Apple in modified form as “stacks”—see Apple Support, “Mac Basics” for an example of how they work.

sort of like a singular personal web page. Steve Jobs took the Mac project over from the fellow who started it, the late Jef Raskin, and soon files appeared.¹⁶

There are folderless versions of operating systems that exist as experimental models, and the demise of the file/folder metaphor for the organization and storage of information is actively discussed on the internet today.

In a post on the question and answer site Quora, a user named Michael Hopkins noted that:

The metaphor has served its purpose. It was absolutely necessary in the command-line-driven computer interface. Only so many file names could be displayed on-screen at once, and there was no ‘scrolling’ the screen as we’re used to now. So a convention was adopted to limit the number of files in each ‘location’ to make navigation easier. The graphic user interface changed this requirement, but still clung to the old metaphor. Some third-party developers tried to create workable 3D interfaces that used such different paradigms as mind-mapping and forest/tree/leaves, but none were particularly successful, mainly due to slow performance and buggy implementations.¹⁷

While not discussing the history of the file system directly, Hopkins’s answer does suggest that particularly savvy users are aware of the history of the file system within their own use of computers and are questioning the utility of it in the future, even if they are not considering the ways in which the file system model has affected their understanding of work or the passage of time. These conversations are essentially cultural outliers, however—in popular parlance as well as popular culture the idea of the file is alive and well.

So if the file/folder metaphor is functionally dead, companies are actively attempting to move users away from it, and users are both aware of the metaphor and of the fact that it no longer has utility, the question becomes why it continues to have such a hold on our collective consciousness—to the point that it is mentally challenging to write or speak about computer systems without using the terms. The reason for this challenge is that the file/folder metaphor encapsulates a theoretical model of storing information online so embedded in our modes of thought that it is a primary lens through which our information intake and output is mediated, and thus has become an abstraction that people do not often think about. However, several unintended modes of thinking come along with that abstracted theoretical model, selected first by IBM and AT&T engineers in the development of computers pre-1980s and then further reinforced by Steve Jobs over the original choices made by Jef Raskin. In fact, Lanier goes on to note that “the file is a set of philosophical ideas made into eternal flesh. The ideas expressed by the file include the notion that human expression comes in severable chunks that can be organized as

¹⁶ Lanier, *You Are Not a Gadget*, 13.

¹⁷ Hopkins, “Dropbox.”

leaves on an abstract tree—and that the chunks have versions and need to be matched to compatible applications.”¹⁸

Such embedded modes of thinking, and the unintended consequences thereof, have their analogues in the non-digital world, perhaps best articulated through Ferdinand de Saussure’s now century-old examination of semiotics and language. For example, in describing the relationship between sound and thought (or more abstractly, between *significant* and *signifié*) in noting the differences between the English *sheep* and *mutton* and the French *mouton*, he states:

“Le français *mouton* peut avoir la même signification que l’anglais *sheep*, mais non la même valeur, et cela pour plusieurs raisons, en particulier parce qu’en parlant d’une pièce de viande apprêtée et servie sur la table, l’anglais dit *mutton* et non *sheep*. La différence de valeur entre *sheep* et *mouton* tient à ce que le premier a à côté de lui un second terme, ce qui n’est pas le cas pour le mot français.”¹⁹

[“the French word *mouton* may have the same meaning as the English word *sheep*; but it does not have the same value. There are various reasons for this, but in particular the fact that the English word for the meat of the animal as prepared and served for a meal, is not *sheep* but *mutton*. The difference in value between *sheep* and *mouton* hinges on the fact that in English there is also another word *mutton* for the meat, whereas *mouton* in French covers both.”]²⁰

Although Saussure’s concern is primarily with language rather than larger questions of thought (his followers, such as Roland Barthes and Jacques Derrida, would be the ones to expand the metaphor to thought in general), this concept still has implications for the development of software platforms for the humanities, especially in grant-funded projects when day-to-day development is undertaken by those without specialist knowledge of the material items displayed and described by these selfsame platforms. When a collaborative team develops a tool, they communicate via the mechanism of language, and in that communication are subject to the split between *significant* and *signifié*. This means that the terminology used to articulate an idea is always understood only imperfectly. At times this imperfect understanding, while unavoidable, will become a hindrance because implicit concepts for the researcher or subject specialist may not be understood in the same way by a developer.

The implications of the division between the implicit understanding of the researcher versus that of the subject specialist can be seen in an anecdote recounted to me by a colleague working on a large scale, multi-person collaborative research project to digitize, transcribe, annotate, and describe medieval manuscript texts. One of the heads of the team on the computational side, when first exposed to the physical objects they had been working with, expressed surprise that they were not all the same size. To

¹⁸ Lanier, *You Are Not a Gadget*, 13.

¹⁹ Saussure, *Cours de linguistique générale*, 160.

²⁰ Saussure, *Course in General Linguistics*, trans. Harris, 114.

those of us who come to the digital humanities through book history, and particularly through codicology, such differences in size are implicitly understood and do not require mention. Form follows function, and thus a Book of Hours will be small, the better to fit in a pocket or purse, while something like the “Codex Gigas” (Kongliga Biblioteket Manuscript A 148), the largest medieval manuscript in the world, is a presentation text nearly a metre tall intended to be read to an audience. Although supposedly only tangential to the actual “content”—the text of the item—such material aspects say things about their portability, their production, and their use. Yet, because the process of digitization and its apparatus of colour bars and rulers was largely incidental to the programmer’s work, the image files created gave the implicit impression that the material objects they described were uniform in size. The materiality of the object, even one as large as the “Codex Gigas,” was abstracted into mere content to be inputted into a black box.

The programmer mentioned above is brilliant, an expert in his particular field, and yet because he experienced these manuscripts in one particular, narrowly focused way, a set of implicit assumptions became part of his modes of thought regarding them. As more and more students and researchers first experience medieval material culture through online media, combatting or counteracting those sets of implied assumptions will be paramount. As Saussure notes, the relationship between signs exists “comme une série de subdivisions contiguës dessinées à la fois sur le plan indéfini des idées confuses (A) et sur celui non moins indéterminé des sons (B)”²¹ [“as a series of adjoining subdivisions simultaneously imprinted both on the plane of vague, amorphous thought (A) and on the equally featureless plane of sound (B)”].²² In other words, how an idea is encoded and expressed to the world in speech or writing is every bit as important as the idea itself.

Of course, for the purposes of discussing the material object semiotics as Saussure describes it has a problem: it is primarily driven by a dyadic representation of language as a medium of communication and knowledge transference, and assumes the sorts of thinking associated with human beings. Computational analysis is limited by the bounds of what the computer can do, as are the mechanisms of online display—and neither of them are particularly good at the concept of so-called “fuzzy,” or multivalent thinking. Moreover, Saussure’s dyad presumes that the sign is “amorphous” or “featureless” and thus does not take into account the actual process of encoding or reception—something that must be considered when working on a digital platform. For this reason, it is worthwhile not to operate solely in the abstracted, linguistically centred world of Saussurean semiotics when thinking about the affordances and limitations of online platforms, but to consider the more cognitively oriented semiotic of Charles Sanders Peirce in conjunction with the framework articulated by John Law in his discussion of Actor-Network Theory.

Peirce’s semiotic theory states that a sign is “anything which is so determined by something else, called its Object, and so determines an effect upon a person, which

21 Saussure, *Cours de linguistique générale*, 155–156.

22 Saussure, *Course in General Linguistics*, 110.

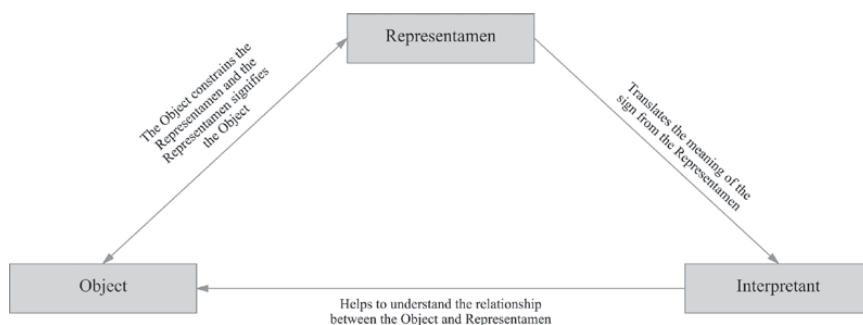


Figure 5.1. The triadic relationship of Peirce's semiotic theory. Note how the arrow between the Object and Representamen is reflective of the relationship between *significant* and *signifié*, but that the addition of the Interpretant provides additional contextual understanding of the sign.

effect I call its Interpretant, that the latter is thereby mediately determined by the former.”²³ That is, a sign is composed of three parts—the representation of something (the Representamen), the restrictions placed on our understanding of that something (the Object), and the understanding that we gain from the connection between the two (the Interpretant). If we map Saussure's semiotics onto this Peircian model, the Representamen would equate to the signifier and the Object to the signified, and in diagram form the relationship between the three can be seen in Figure 5.1.

Peirce never completed a definitive explanation of his theory of signs; what we have is a number of works in progress and correspondence, published after his death, which deal with the relationship between the three categories in slightly different ways. All of them have the same triadic structure, but each of them treats the relationship between the elements slightly differently.

In the earliest account, Peirce believes that the Interpretant acts as a more developed version of the relationship between the Representamen and the Object. The chain of connections this generates is similar to the two planes Saussure mentions, but handled dynamically instead of statically as each link in the chain is generated by the one that came before it. In this early account Peirce makes a connection between the words “homme” and “man” that is reminiscent of Saussure's sheep/mouton/mutton example. In describing the Interpretant, he suggests that the reader

suppose we look out the word *homme* in a French dictionary; we shall find opposite to it the word *man*, which, so placed, represents *homme* as representing the same two-legged creature which *man* itself represents. By a further accumulation of instances, it would be found that every comparison requires, besides the related thing, the ground and the correlate, also a mediating representation which represents the relate to be a representation of the same correlate which this mediating representation itself represents. Such a

²³ Peirce, *The Essential Peirce*, 478.

mediating representation may be termed an interpretant, because it fulfills the office of an interpreter, who says that a foreigner says the same thing which he himself says.²⁴

Peirce considers a ground “a pure abstraction” that, when referred to, “constitutes a quality or general attribute ... reference to a ground cannot be prescinded from being, but being can be prescinded from it.” Likewise, a correlate is the quality that is only understood “by means of its contrast with or similarity to another.”²⁵ Rather than suggesting that the two linguistic environments are separate planes, Peirce is suggesting that in translating “homme” to “man” the Interpretant does the work of connecting whatever is expressed by the relationship between the two to our understanding. Moreover, in doing so it makes it clear what exactly is being expressed by that relationship. Similarly a visual sign, such as the universal signs for man and woman found on bathroom doors, can be seen as a separate environment from the real, lived reality of gender in our daily lives. It is our ability to generate a correlate between the abstracted icon and the real people around us—informed in turn by socialization and cultural expectations—that lets us immediately assume that the man and woman icons relate to those individuals who identify in turn as men or women themselves. Likewise, our ability to translate what is essentially a stack of unrelated image files into a mental representation of a manuscript or printed codex relies on our ability to create such correlates as mediating representations between the images and the physical object.

Thus, the Interpretant acts as a lens for the recipient of that sign, as the Interpretant of the Representamen/Object dyad will be different for each individual. This means that there are effectively at least two layers of abstraction—the “natural” layer of abstraction that is the result of the reader interacting with the material artefact (the reader acting as Interpretant of the Representamen/Object dyad) and the imposed, mediating layer of abstraction generated by the online platform and the decisions made in rendering the material object online. The choices made in presentation for that online platform could differ greatly from the information expressed by the materiality of the manuscript itself, depending on the developer or development team’s understanding of what is being expressed—in the anecdotal example, the understanding of the size of medieval codices as a function of the presentation of information within them.

Because a virtual representation of a material artefact—even one which is collecting currently divided pieces and reconstituting the “original”—is always a mediated work it should not be considered a facsimile as we typically understand it, but rather as an adaptation of that work—a link in an infinite semiotic chain. Infinite semiosis, as Peirce describes it, means that “anything which determines something else (its Interpretant) to refer an object to which itself refers (its Object) in the same way, the Interpretant becoming in turn a sign, and so on ad infinitum.”²⁶ Sign here being taken to be equivalent

²⁴ Peirce, *Writings of Charles Sanders Peirce: A Chronological Edition*. vol. 2, 53–54.

²⁵ Peirce, *Writings of Charles Sanders Peirce*, 53.

²⁶ Peirce, *Collected Papers*, vol. 2, 169.

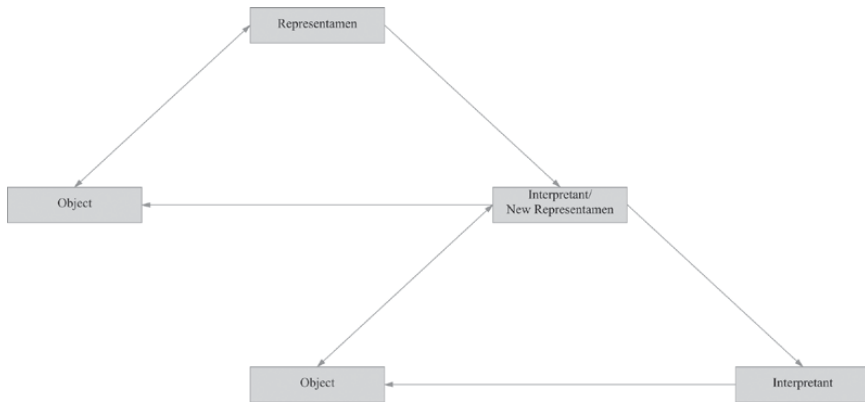


Figure 5.2. Infinite Semiosis—the Interpretant of the first sign becomes the Representamen of the next sign in the chain.

to Representamen, it becomes a part of a new triad, and the Interpretant of that triad is itself a more fully developed version of that original sign.

The major problem with the concept of a chain of signs, of course, is that when it ends, there is no further Interpretant generated. Once that occurs, the result cannot be considered a sign anymore. Peirce suggests as much when he states that “if the series of successive interpretants comes to an end, the sign is thereby rendered imperfect, at least.”²⁷ However, even if this ending were to occur, it is only possible if we assume that ideas exist in a closed system. Since ideas continue to evolve over time, the system is never truly closed and ideas regarding the sign—or in this case, the interpretation of the manuscript either in its material or digital instantiations—continue to evolve as well.

Peirce’s supposition that signs are “imperfect” originates out of the idea that infinite semiosis occurs within a single individual’s mind. In that model, only the output is available to the audience. This has similarities to the technological black box already discussed: something that receives input and generates output, but whose operations are opaque to the viewer. However, when dealing with inscribed works, the output is not simply the product of thinking, but is part of a process of inscription, reception, and re-inscription that is continually occurring. Instead, even if we assume an individual’s thinking about the “content” of the material object is the output of the black box, expressed through the mechanism of a technological platform, that output itself becomes new input for other semiotic chains. As such the chain of infinite semiosis operates between individuals as well as within each individual’s mind, and might better be considered as a *network* of competing and cooperating interpretations rather than as a linear chain. Actor-Network Theory can thus help to explain some of the features of this network and how the ideas encoded in the material artefact both mediate and are mediated by the means of its production and reception.

²⁷ Peirce, *Collected Papers*, vol. 2, 169–70.

Actor-Network Theory and the Platform

As John Law envisions it,

Actor network theory is a ruthless application of *semiotics*. It tells that entities take their form and acquire their attributes as a result of their relations with other entities. In this scheme of things entities have no inherent qualities: essentialist divisions are thrown on the bonfire of the dualisms. ... it is not, in this semiotic world-view, that there are no divisions. It is rather that such division or distinctions are understood as effects or outcomes. They are not given in the order of things.²⁸

This emphasis he makes on effects or outcomes, rather than essentialist divisions, means that what we see as singular actors are really the action of “patterned networks of diverse (not simply human) materials.”²⁹ To explain this, he begins from the standpoint of sociology of science, where it is argued that “knowledge is a social product rather than something generated by through [*sic*] the operation of a privileged scientific method.”³⁰ Knowledge, in this case, is not knowledge as we generally conceive of it, as a purely

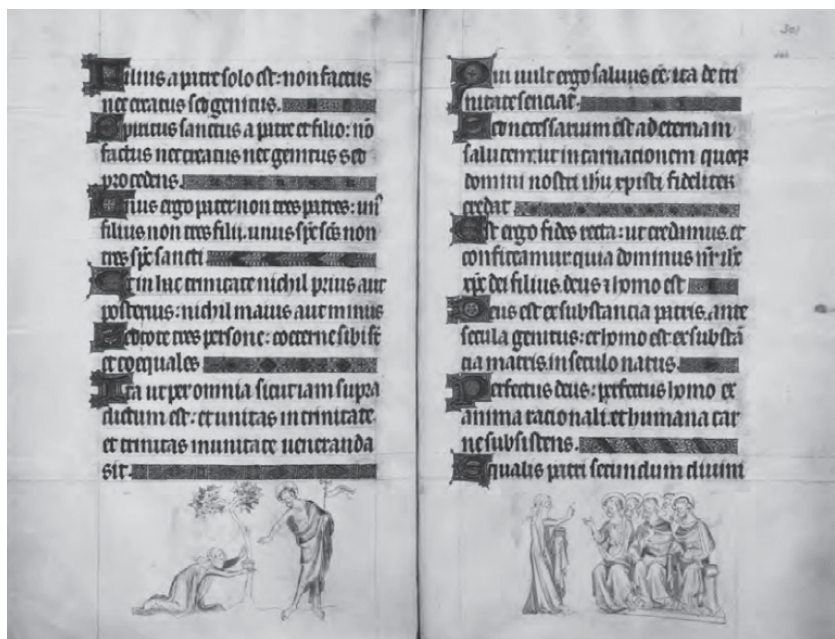


Figure 5.3. A constructed image intended to represent the opening showing fol. 300v-301r of the “Queen Mary Psalter.”

28 Law, “After ANT,” 3.

29 Law, “Notes,” 2.

30 Law, “Notes,” 2.

Platonic set of concepts, facts, and intuitions. Rather, knowledge is embodied in material forms, as the product of work wherein the materials and ideas that are used to produce that knowledge are organized into a network that overcomes the natural desire of these individual pieces to “make off on their own,” or become disassociated from each other.³¹ This network extends out into the world through the tools and ideas the researcher is exposed to, and that conceptual “footprint” serves as the means of mediation between the researcher and the world at large. In effect, the researcher is not just a single individual with a mind divorced from the concerns of the body, but a part of a network of overlapping materials and ideas receiving, processing, and encoding ideas—a single link in a grand semiotic chain.

The effect of tools on how information is received and processed, and how easily such a network of materials and ideas can be obscured if one is not careful in articulating them, can be shown in the example in Figure 5.3, an image from the “Queen Mary Psalter” (Royal MS 2 B VII, Fol. 300v–301r).³² The image as presented there looks as though it is two facing pages of the physical codex book. However, the British Library presents the two pages as depicted in Figure 5.4 when you look at it in a web browser.³³

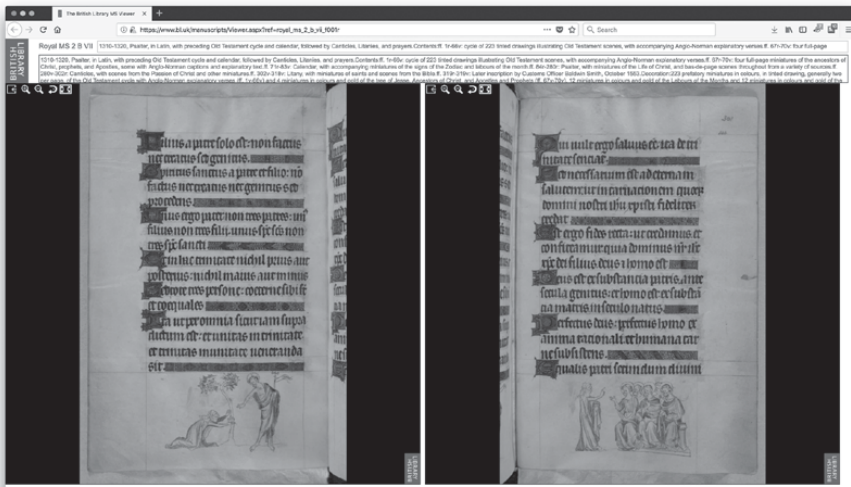


Figure 5.4. Fol. 300v and 301r as they appear on the British Library's Digitised Manuscripts site.

31 Law, “Notes,” 2.

32 British Library, Royal MS 2 B. VII.

33 British Library, Royal MS 2 B. VII. www.bl.uk/manuscripts/Viewer.aspx?ref=royal_ms_2_b_vii_f300v.

Several steps are necessary to make the pages appear as they do in the first image. I am required to capture each page as a separate image file, then merge the image files in a way that places them next to each other, which produces the effect of the particular opening in the book. In doing so, I utilize my computer, a web browser, a piece of graphics software, and the ability of my computer to take an image and insert it into the document I am working on. This does not even take into account the network of camera equipment, servers, and skilled workers that were necessary to digitize and store the successive pages in the first place. Or, for that matter, the political network that resulted in the acquisition of the manuscript by Queen Mary, thus allowing for its eventual deposit in the British Library from the royal household, the economic network that allowed for the acquisition of the text, or in networks of scribal production, manuscript acquisition, and bindery that went into the development of the original psalter. All of that work, from the point at which the first piece of parchment was written upon to the point that the image was placed in this chapter, is functionally invisible. Instead, we perceive the placement of the images from the “Queen Mary Psalter” as a single, unitary thing—a single block, as Law puts it. And, as Law notes, “if a network acts as a single block, then it disappears, to be replaced by the action itself and the seemingly simple author of that action.”³⁴

When a researcher does not consider a tool for presentation part of their theoretical process, they are in effect doing the same work that I have done when putting the two images together—rendering the entire apparatus necessary to put the virtual facsimile of the text online, with all of the decisions made in doing so, to a single block with a single author which, in turn, is dedicated to a single action—the conveyance of “content” to the waiting user. In effect, the adaptation is being presented as more than functionally similar to the original—it is being presented as though it *is* the original, with only the medium of presentation changed. If we stop and think about it, this is an impossibility, yet it is something we cognitively do on a daily basis when approaching these large-scale, mediating networks. And, as the anecdote about my colleague’s team member shows, making those assumptions about material artefacts can at times result in errors in judgement. In the case of the programmer, the error was relatively mild, but as our use of digital technologies in the practice of humanities scholarship grows—especially with the use of distant reading and other “Big Data” methodologies becoming more and more common—such an error can have grave consequences that could go unnoticed because the actual objects have been abstracted beyond our ability to return to them for a needed corrective. Instead of assuming that the technology exists independent of the content driven by that technology, we must as part of our scholarly approach remain aware of the network and the assumptions made by it. Rather than such collapse of the illusion of a singular object occurring by happenstance, however, I believe that “conscious,” or careful and reasoned, collapse can be achieved through a constant questioning of the network and the pieces that comprise it. This action I refer to as “radical transparency”

34 Law, “Notes,” 380.

in the development of tools and methodologies. Such questioned networks, and the transparency they generate, will allow us to both acknowledge that our work with digital and material tools is itself the result of our status as an Interpretant of the Object/Representamen dyad and to make our status as part of the semiotic network readily apparent to the scholars and interested individuals who will be the eventual audience for the products of our study.

The need for radical transparency as a methodology in the development of the platform as a theoretical abstraction—whether as the printed or manuscript book or as text on a computer screen—means we have to directly acknowledge that it takes its attributes from the tools that currently exist at the time of its composition. In the RCA advertisement the metaphor of the file and folder became an easy way to get the concept of computer storage across to the mass public, aided by the fact that the mechanism of storage is functionally akin to a file when placed into a filesystem. However, because the theoretical basis of that abstraction was not firmly articulated (who wants to read a manifesto in their ad copy, after all) its basis as a metaphor was lost. Instead, over time the concept of the file broadened its semantic scope, becoming synonymous not just with the physical place where the information is stored but with the information itself. The attendant loss of the quotation marks and the eventual broadening of semantic scope of the word “file” to include not just the physical means of storage but the information itself are thus markers of the increasing comfort users had with the file/folder metaphor and the underlying technology, but hidden within that comfort level are, as Lanier notes, serious choices about how we express and process information. Until the network surrounding the metaphor is itself considered, as the Quora poster is beginning to do, those choices influence our development of further tools and methods without our knowledge.

While I am aware that the brief examination of the history of the file/folder metaphor above may have seemed digressive at first, within the context of the theoretical apparatus of Infinite Semiosis and the Actor-Network the larger point must be understood that the decisions we make—even so innocuous a decision as the choice to call the information storage mechanism a “file” in a piece of ad copy—have both long and short-term implications for the life of a digital humanities project. Chief amongst these is the differences in the ways professional software developers and the academic community treat the presentation of information.

As one O'Reilly manual on software development, designed for the professional market, states “Software development is all about change, and moving to your next iteration is no exception You've got to ... adjust your stories and expectations based on what the customer wants *NOW*, not a month ago.”³⁵ Taking this logic at face value, it is clear that neither software development's iterative cycle, nor the books and bootcamps

35 Miles and Pilone, *Head First Software Development*. Also note how the language of Miles and Pilone reinforces, but does not really consider critically, the notion of “update” that Chun refers to in *Updating to Remain the Same*.

increasingly responsible for training programmers for the market, are concerned primarily with recording the process of development or the interaction between the material artefact and its virtual adaptation. Instead, the primary concern is with speed and profit. The responsibility of the software developer is to present the customer with a tool, application, or platform that meets the definitions of “what the customer wants,” usually articulated in the form of a Solution Design Document.

Such a document lays out the expectations of the client for the tool or platform and prevents what software engineers call “scope creep”: the slowly shifting sets of expectations that add time and expense to any software project. What they do not do, though, is suggest that the developer must have the sort of deep subject knowledge that scholars possess on the subject. Thus, the developer’s idea of “what the customer wants” is governed by their own sets of ideas and expectations, which as mentioned above privilege change and speed over articulation of the theoretical decisions made in presenting the material artefact online. In fact, it is a truism of the software development industry that developers hate to document how their code actually functions, relegating that to a group of skilled technical writers who do not necessarily have access to the thought process of the developer when writing the code and have no interest in articulating anything but how the software functions as a product. Moreover, due to the desire to complete the tool quickly, under budget, and in a way that works with other tools, solutions are often built as “one size fits all” models. Such models do not always take into account edge cases—the liminal spaces where so much productive work in scholarship is undertaken.

The language and models of software development are thus concerned primarily with the creation of the tool as an appliance to be used, rather than educating the user about the tool’s development in a way that makes them understand the theoretical and practical issues that went into its production. They are concerned with the short-term development cycle, intended to maximize profit, rather than with long-term consideration of the effect of those tools on cultural and scholarly development. And most scholars are ill-equipped to notice that these built-in assumptions on the part of technologists and programmers are occurring. Nevertheless, while it is unrealistic to ask most humanities scholars to become as well-versed in the tools and methods of software development as a professional programmer, it is our responsibility as scholars, researchers, and instructors of often unique medieval works to interrogate the assumptions programmers make, remain aware of the implications of these assumptions, and to push back against design choices when they might have negative, and costly, impacts on our understanding of the material artefact as cultural heritage item.

A Nod Towards “Radical Transparency”: Developing the Minor Works of John Lydgate Virtual Archive

One of the “strengths” of the digital humanities, as articulated in numerous articles, web posts, chapters, and conference presentations, is the fact that the work of producing digital projects is often collaborative in nature. It requires both the work of the individual researcher with deep subject knowledge and the work of programmers, metadata

librarians, and other specialists who are well-versed in their own areas of expertise, but may not have that same deep subject knowledge. However, I qualify “strength” in the initial sentence of this paragraph because while collaboration in itself is not an issue—it makes logical sense for the various parties involved in a project to work to their particular strengths—as the previous section notes the strength of collaboration as a method of rapidly instantiating a digital platform can become an issue philosophically.³⁶ Often, and quite subtly, the needs of that platform begin to shape the direction of the scholarly enterprise, rather than the other way around—the user starts to become the machine, to use Chun’s formulation.

In some ways, this is an unavoidable result of the influence of the network on the process of re-inscription. Since the network doing the work of reading, interpretation, and re-inscription is always different to that which initially created the item, part of the work of developing a tool has to be articulation of the theoretical basis for the tool and the methodology planned to both do the work and to present the result. Until a viewer of an online platform has some sense of the network, it remains functionally equivalent to a black box, and thus cannot be fully taken into account when using the tool in the process of scholarship.

For example, the terms “database” and “spreadsheet” are often used interchangeably by academics who have project-related data they want to incorporate into a finished product. Functionally, they even *appear* to be the same at first glance—a table of cells that information is placed into. A database, however, has further information relating multiple tables to each other via one or more indices and is thus designed to be rapidly searchable along many axes, while a spreadsheet is often limited to a smaller number of sheets, searchable only along a few axes. This lack of understanding of the difference can cause problems and delays when a developer asks for a record not only of all the information in the database, but of its structure, and the researcher provides the spreadsheets they have been using. To a human being, these are functionally similar, because a human being can do the process of evaluation and comparison without explicit instructions. A computer, on the other hand, cannot.

For these reasons, taking a look at the larger issues at play in the development of a simple tool can help us to understand the intersection between the technical, the theoretical, and the methodological, and how they all work together in the presentation of cultural heritage items online. Understanding that intersection, in turn, helps to avoid the obfuscation or elimination of *context* and *paratext* when presenting these items online.

“Context” and “paratext,” here, are set against the term “content,” referred to above, rather than to “text” as they are in most relationships between the words. As I use the term, the “content” of an item is the information it contains divorced from its method of presentation, and is largely analogous to the “text” as articulated by scholars but not exactly representative of it. The paratext, conversely, consists of all of those “extra” bits

36 Indeed, as of November 16, 2017 a simple Google Search for “digital humanities collaboration” provides over four million results. Limiting it by “strengths” returns approximately 350,000 results, while limiting it by “weaknesses” provides 289,000 results.

that make the notion of a truly fluid text difficult—introductions, illustrations, notes written in the margins, and so forth. Finally, the “context” is all of the ancillary elements, such as the historical situation it was written within, the limitations of the tools used, the mindset of the original author or subsequent scribes, and the understanding of the reader in approaching the material. Within a digital cultural heritage project one element, typically the content, is usually foregrounded, but even when not recognized all of these elements exist as palimpsests that are only imperfectly erased from or elided in the digital adaptation. Part of my thinking in developing the Minor Works of John Lydgate (hereafter, in the spirit of most digital humanities projects, referred to as MWJL) is to point out these easily overlooked elements.

MWJL has, as its goal, the collection and presentation of works by the fifteenth-century English poet John Lydgate. Lydgate was an important figure in Plantagenet court culture and politics between the first decade of the fifteenth century and his death in 1449 or 1450, and in literary circles 150 years thereafter. These works thus inform our understanding of religious and dramatic practice in the period just prior to the English Reformation, but are not often anthologized or discussed outside of academic circles. The primary goal of the website is to help correct this by presenting transcriptions of the Lydgate works alongside images taken from the manuscripts that contain them. However, a secondary goal of the site is to use three-dimensional and other models to articulate the relationship between the text, paratext, and context of the poems, an aspect currently missing from digital archives of most medieval English works.

Although a writer of both dramatic and poetic works, Lydgate is often most remembered for the former. As Claire Sponsler has pointed out, this is an artefact of the association between the poetic form and “literature,” which has significant impacts on our understanding of English culture in the Middle Ages, and is itself an artefact of the inscription and re-inscription of ideas about literary worth by contextual networks over time.³⁷ It has also worked to the detriment of our understanding of medieval poetry and drama in general and Lydgate’s works in particular. Additionally, the manuscript object provides material contexts that can provide clues to its intended audience, the historical and social situations at the time of its production, and its original as well as later purposes. For these reasons, displaying both the text of the object and its contexts is paramount in presenting the object as a virtual facsimile online.

Besides the commentary on the relationship between the network and the material and digital texts that the site provides, it is also necessary because without such work, aspects of English culture are in danger of being lost. For example, unique versions of Lydgate’s *Testament* and “Quis Dabit Capiti Meo Fontem Lacrimarum” (also known as “The Lamentation of Our Lady Maria”) exist in the chantry chapel of the Clopton family at the parish church of the Holy Trinity in Long Melford, Suffolk. These texts are not pages in a manuscript, but instead rendered in carved wood and flaking paint. The poems at Long Melford are not the versions found in academic and student editions, however. Instead, they are versions altered specifically to fit the purpose and space for

37 Sponsler, *The Queen’s Dumbshows*, 1–12.

which they were produced in the chapel, and must be read through interacting with that space.³⁸ They are unique cultural items, as are all of the manuscripts of the 185 known or suspected poems by Lydgate. Yet they are not treated as viable items for study in their own right. This causes that unique nature, and what it says about the cultural history of England, to be subsumed under a single, standard version, with implications for future study of the poet and the semiotic network of understanding surrounding him. In short, the process of inscription, reception, and re-inscription that is constantly ongoing as part of the network is being influenced, however unintentionally, by traditional editorial practice. One of the goals of the site is to open up this black box and make scholars and students aware not only of alternate versions of Lydgate's works but of the ways in which the various witnesses are presented.

I began work on the site by doing an environmental scan of the items either definitely or purportedly written by Lydgate in the Digital Index of Middle English Verse (DIMEV), organizing them into a Structured Query Language (or SQL) database based on the connection between particular works and particular manuscript witnesses. After placing all the items in the database, I was able to quickly determine that, of the original manuscripts and editions of Lydgate in DIMEV:

- 43% of the works are unavailable in print or online editions
- 63% of the unique versions of these poems are unavailable in print or online editions
- 59% were produced over 100 years ago
- 91% were produced over 50 years ago
- Only 1% were produced since the year 2000.

Thus, the majority of scholars working with Lydgate's poems are only working with a few witnesses that were largely produced in the first decades of the twentieth century. While these printed texts are obviously still useful, they often rely on methodologies and assumptions uninformed by recent technological innovation, and thus make assumptions (or have assumptions made about them) that may no longer be valid because they only imperfectly fit into the network of tools and ideas currently at play in medieval scholarship and manuscript studies. MWJL is an attempt to rectify this by increasing the availability of transcriptions alongside manuscript images in a way that is free, open to the public, and in keeping with the best practices for manuscript transcription and description.

If we consider the distinction between the network as I have articulated it earlier and the popular notion of "content" as ideas Platonically divorced from their presentation, the need for this rectification becomes evident. In reality no single text—whether print, manuscript, or digital—exists in an ideal version, as the notion of the semiotic network and infinite semiosis underscores. Each iteration of that text includes shifts in presentation, editorial or reader notes, and in some cases omissions or additions. The version of

38 For more on the particular versions of Lydgate's work at Long Melford in relationship to the architectural space, see Davis, "Lydgate at Long Melford."

39 The full schema and data for this database is available at www.minorworksoflydgate.net/witness_lookup.sql

a particular moment, then, is really just one of a number that compete with each other for our attention, a notion obscured by the prioritization of content in the development of the authoritative edition for mass consumption. Such editions foreground a scholar or scholars' conception of the idealized text, both in form and presentation, over any other versions available, and when presented digitally are also informed by the work of the technical staff that help to make the online site a reality. As a means of rectifying this conceptual bottleneck, a touchstone of development for MWJL is to examine the full process of production and reception of several of Lydgate's poems that are unavailable to most of the public except in single, authoritative texts along with careful delineation of the decision-making process involved in presenting them online.

After completing my scan of the various witnesses of Lydgate, and realizing how much of his work exists in largely unavailable forms, I wrote down a set of design principles to work from while developing the site:⁴⁰

1. Defer to the material object whenever possible. Undue abstraction will be avoided.
2. Follow coding standards, but do not allow the technical standards to obscure the material item and its own unique narrative. When it is necessary to deviate from those standards, explain it by providing guideposts whenever possible.
3. Use the most efficient tools possible. Do not choose a platform simply because it is newer.
4. Visualizations or texts should not be displayed without the underlying data and an explanation of what exactly it means. Someone who does not have a technical background needs to be able to follow both the scholarly and technical explanations. Examples are to be used whenever possible.
5. Respect the audience. Act as a bridge between them and the work, so that what came before is not lost in chasing what is to come.

While not expressed in formal theoretical jargon, when combined with my own personal network of influences they provide a much-needed reference for both the more formal elements of the site—the transcriptions, descriptions, and ancillary elements we would consider paratext in a more formal edition—and the more colloquial elements intended for a mixed audience of scholars, students, and the general public. Any site interested in presenting both the process and the product of their work should make available a similar set of principles, based on the intentions of the site and the theoretical and methodological preferences of the development team.

From this set of overarching principles, I then approached methodology, breaking down the process of development into six concrete steps that would be undertaken at least once for each poem I intended to work with:

1. Acquisition of appropriate image files from the holding institutions.
2. Transcription of each witness of the poem.

⁴⁰ The original, more colloquially written set of guidelines can be seen at Davis, "About the Archive," under the "What are your editorial principles" header.

3. Complete codicological description of each manuscript.
4. Encoding of the transcription and description of each manuscript.
5. Display of the resulting files online.
6. Development of ancillary features that expand on what is possible in a print edition.

These broad steps were broken down still further into manageable pieces, and the site itself was used both as a way to inform an audience where exactly in the process each witness was and as a way to discuss what still needed to be developed in terms of the ancillary features not already on the site.⁴¹ Thus, while it would be reductive to go into detail over every single step in the process, it might be worthwhile to discuss in brief two aspects of the site in greater detail: the philosophy behind the particular version of data encoding used on the site and the process of transcription, as well as a visualization designed to quickly access the corpus of Lydgate texts in an easily understood manner.

Transcription Philosophy and Method

As is evident from the theoretical overlay at the beginning of this chapter and the section outlining my design principles I am interested in presenting the material artefact as I best can in a digital platform, while remaining cognizant that there are going to be elements both gained and lost in doing so. The standard for digital transcription in my field is a particular schema for the eXtensible Markup Language created by the Text Encoding Initiative (XML and TEI, respectively).⁴² That schema offers two methods for the transcription of texts. The first, most common, method places all the textual information and its associated tags under a single `<text>` element within a hierarchical tree structure.⁴³ Note that the very process of creating the framework for displaying the text online makes assumptions privileging the content of the item over its physical characteristics. In fact, these characteristics are only imperfectly captured under the `<physDesc>` element and the elements that exist underneath it. Those elements, in turn, presume that the object being described will be a paper or parchment page in a codex book, which will not work for an example such as the Clopton chapel I referenced earlier in this chapter. Thus, design decisions made by the TEI Consortium in the development of a single, overarching standard for the display of both manuscript and printed text already leaves out important elements of Lydgate's textual output.

There is a solution, however. The most recent revision of the TEI standard introduces a new head element, `<sourceDoc>`, which is intended to capture both the transcription and the physical aspects of a single document.⁴⁴ Rather than separating out text and paratext `<sourceDoc>` provides the option to create what the consortium calls an

⁴¹ Davis, "Works."

⁴² The two terms are often used interchangeably, but really the TEI schema is a "namespace," or organizational framework, that uses XML as its platform of choice. See further "Extensible Markup Language (XML)" and "Namespaces in XML 1.0," respectively.

⁴³ TEI: Text Encoding Initiative. "4. Default Text Structure"

⁴⁴ TEI: Text Encoding Initiative. "TEI element sourceDoc."

Figure 5.6. The <sourceDoc>-based encoding for the version of Lydgate's "Quid Dabit Meo Capiti Fontem Lacrimarum" in Jesus College Q.G.8, Jesus College, Cambridge, as rendered via the oXygen XML Editor.

Armed with the affordances created by this new head element in TEI, I begin each transcription by writing the characters, as I see them, into a word processing document. While this differs from the method usually taught, based on writing each of the characters into a notebook by hand so as to capture the nuances of the scribal hand, the increasing likelihood that an archive will allow researchers to take their own digital photographs of a manuscript for reference makes it easier to type the characters out and then refer to the reference shots in cases where there are scribal oddities.⁴⁶ This also alleviates any chance of a malformed character as a result of imperfect copying.

Because both Microsoft Word and Libre/OpenOffice store their documents as wrapped XML files, I then unzip the file and utilize an XSL transformation to render the underlying XML code—WordProcessingML for Microsoft Word and OpenDocument for Libre/

46 For a description of the more traditional method, see Brown, *A Guide to Western Historical Scripts from Antiquity to 1600* as well as Clemens and Graham, *Introduction to Manuscript Studies*.

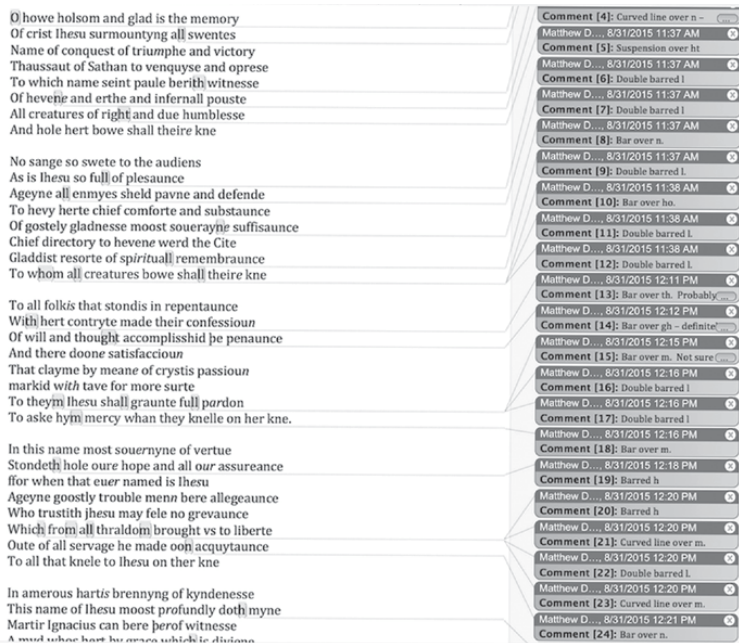


Figure 5.7. The author's master transcription file for the witness of the *Testament of John Lydgate* in Huntington HM140.

OpenOffice—into a format that is compliant with TEI's schema.⁴⁷ From there, the resulting information is manually wrapped with the header information for a TEI file and those elements that cannot automatically be transformed—the notes and other ancillary features captured via the word processor's commenting feature—are added to the resulting file manually.⁴⁸ During this time the file is also checked against the images, and if everything is correct, a second XSL transformation is run to render the file into static HTML.

Note that this process, especially the rendering of the files into static forms whenever possible, harkens back to the basic design philosophy outlined above. HTML is ubiquitous, and as such rendering the file into HTML makes it much more likely that any future web browsers will be able to understand and process the files without further development work. Those features that cannot be handled statically, such as line comparisons or the three-dimensional model of the Clopton chapel that augments the existing image/transcription display on the pages, are handled via JavaScript. Moreover, that model is presented on the site in a way that deliberately makes it clear that it is incomplete, and

⁴⁷ On .docx files as wrapped XML, see Jones, "Intro to Word XML Part 1."

⁴⁸ The XSL file used to do the initial transformation of the Word XML can be found at www.minorworksoflydgate.net/XML/XQuery/xsl_word_sourceDoc.xsl. The document that does the work of final transformation for upload to the site can be found at www.minorworksoflydgate.net/XML/XQuery/chunker.xsl.

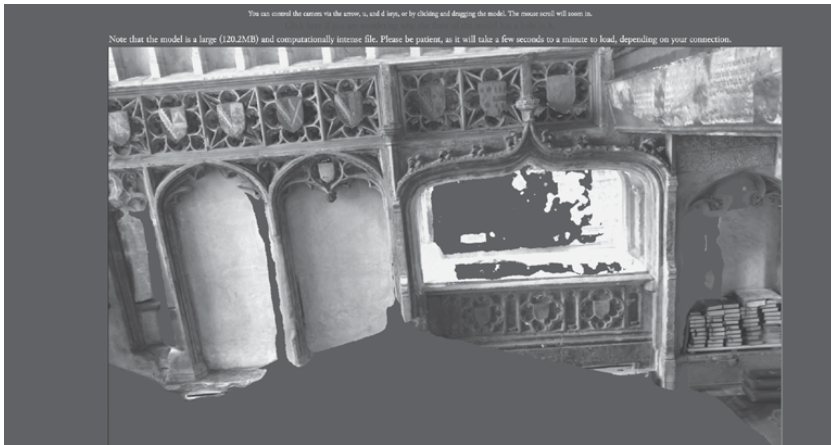


Figure 5.8. Model of the Clopton chantry chapel, Holy Trinity, Long Melford. Note the use of empty space in the tomb opening, for example, to remind a viewer that this is a model of a real, physical space.⁴⁹

so forces a viewer to recognize that their experience of the chapel space is in fact mediated. This, in turn, forces recognition of the network underlying the development of the text for display online and makes a viewer consider the physicality of the space in a way that might not otherwise occur.

JavaScript is also used to display the relationships between aspects of these texts as they are defined in the database created as part of the initial work on the site. This gives users of the site access to the underlying network surrounding these texts as they exist today and makes them aware of holes that may exist in current scholarship. Furthermore, it allows them to, at a glance, see what texts exist alongside each other in the manuscript witnesses, where those manuscripts are located currently, and who was involved in their production and distribution.

Both the three-dimensional model and the force-directed graph are admittedly crude approximations of the physical space at Long Melford and the relational network associated with existing Lydgate holdings, respectively. However, they serve their purpose: to remind readers of the physicality of these material artefacts and the fact that they do not exist as simple images on a website or Google Image Search, but instead have a real and ongoing life that the digital version can only approximate.

Conclusion

While by no means a comprehensive discussion of the theoretical and methodological decisions made in the development of MWJL, I hope that the discussion of my

⁴⁹ Davis, "Clopton Chantry Capel." The model was developed using photogrammetry and is displayed using the Three.js JavaScript library.

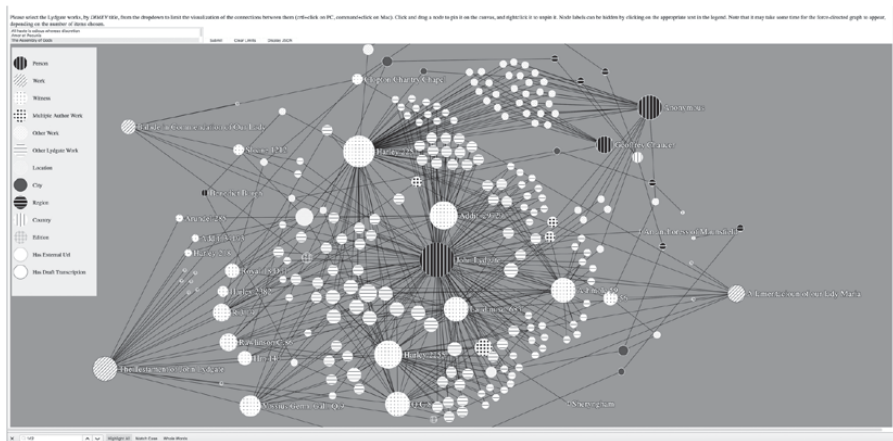


Figure 5.9. Force-directed model indicating the relationship between the three texts at Holy Trinity, Long Melford, in all witnesses.⁵⁰

transcription method and the development of two aspects of the site illustrate how clear, careful, and transparent articulation of the methodological as well as theoretical approach to a project can help to illuminate the resulting virtual product and counteract the sort of digital presentism Chun notes and Parker's commentary reinforces. More importantly, the examination of the issues surrounding the semiotic network, as illustrated by the "Queen Mary Psalter" and the file/folder metaphor, indicate that requiring transparency and a continual awareness, repeatedly articulated, of the network surrounding any digital or analogue tool or underlying idea is paramount. Doing otherwise obscures much of the work of building and revealing culture, and furthermore skews our understanding of what we are actually experiencing when we use that tool or re-inscribe that idea in our own work.

To explain how, why, and with what method a text is transcribed, for example, is important, but equally important are the decisions made in the development of encoding standards. The TEI's decision to base their architecture on the codex book, while sound in most cases, would have had grave consequences for our understanding of the Clopton chapel, and it is only with the recent revision of the TEI standard that it is possible to even approximately display the chapel as it is *in situ* while staying compliant with the standard as written. The ability to solve that problem through a combination of the <sourceDoc>-based encoding methodology and ancillary tools such as the three-dimensional model and force-directed chart underscores that the transcribed text in both its encoded form and its virtual presentation is an adaptation of what came before. That adaptation is profoundly reliant on networks of production and influence

⁵⁰ Davis, "Force-Directed Graph." A version of this model with all the works of Lydgate in the database can also be seen if you click on the "Clear Limits" button. It was created using the d3.js JavaScript library.

that are largely obscured, but which have the potential to fundamentally change our understanding of these cultural heritage items. Scholars must be diligent in making clear the decisions they have made in preparing these items for display online and how those decisions both support and detract from the affordances of the material original.

Since all editions and facsimiles are ultimately adaptations of a changing and mutable set of ideas, a virtual adaptation of a unique cultural object should attempt to accurately model the material original, but acknowledge that it is not itself an adequate substitution for that material original. Rather than claim a fidelity that does not exist, scholars should firmly and transparently articulate both the theoretical and methodological stakes of the project and the ways in which they approach the text and its presentation online. This work is already largely done when it comes to transcription methods and the display of content, but falls woefully behind when it comes to context and paratext. In articulating the decisions made in choosing or developing a platform and the methods used in presentation of content a reader is at least made aware of the infrastructural palimpsests incorporated in the online presentation of that content. In describing my work with MWJL I have only articulated one method of doing this work, and there are obviously others, but ultimately both theory and praxis must be articulated for a digital tool to have its greatest use.

Bibliography

- Allen, Mike. "Sean Parker unloads on Facebook 'exploiting' human psychology." Axios.com. www.axios.com/sean-parker-unloads-on-facebook-2508036343.html.
- Apple Support. "Mac Basics: Stacks are your shortcut in the Dock." March 23, 2016. <https://support.apple.com/en-us/HT201955>.
- Brown, Michelle. *A Guide to Western Historical Scripts from Antiquity to 1600*. Toronto: University of Toronto Press, 1990.
- Carruthers, Mary J. *The Book of Memory: A Study in Memory in Medieval Culture*. Cambridge: Cambridge University Press, 1990.
- Chun, Wendy Hui Kyoung. *Updating to Remain the Same: Habitual New Media*. Cambridge, MA: MIT Press, 2016.
- Clemens, Raymond, and Timothy Graham. *Introduction to Manuscript Studies*. Ithaca: Cornell University Press, 2007.
- Davis, Matthew Evan. "Lydgate at Long Melford: Reassessing the *Testament* and 'Quis Dabit Meo Capiti Fontem Lacrimarum' in Their Local Context." *The Journal of Medieval Religious Cultures* 43 (2017): 77–114.
- . *The Minor Works of John Lydgate*. www.minorworksoflydgate.net.
- . "About the Archive." www.minorworksoflydgate.net/archive.html.
- . "Force-Directed Graph." www.minorworksoflydgate.net/forcedirect.php?select%5B%5D=4&select%5B%5D=83&select%5B%5D=163.
- . "Clopton Chantry Chapel." www.minorworksoflydgate.net/Model/three/examples/chantry_chapel.html.
- . "Works." www.minorworksoflydgate.net/works.html.
- Drout, Michael D. C. *How Tradition Works*. Tempe: Arizona Center for Medieval and Renaissance Studies, 2006.

- "Extensible Markup Language (XML)." W3C. www.w3.org/XML/.
- Gibbs, Fred, and Trevor Owens. "Building Better Digital Humanities Tools: Toward Broader Audiences and User-centered Designs." *DHQ: Digital Humanities Quarterly* 6:2. www.digitalhumanities.org/dhq/vol/6/2/000136/000136.html.
- Glass, Grant. "The Black Box." HASTAC. www.hastac.org/blogs/grantglass/2015/10/14/black-box.
- Hopkins, Michael. Comment, "Dropbox (product): Will the metaphor of files and folders become obsolete?" October 23, 2011. www.quora.com/Dropbox-product-Will-the-metaphor-of-files-and-folders-become-obsolete
- Jones, Brian. "Intro to Word XML Part 1: Simple Word Document" in *Brian Jones: Office Solutions*. https://blogs.msdn.microsoft.com/brian_jones/2005/07/05/intro-to-word-xml-part-1-simple-word-document/.
- Knappett, Carl. "Networks of Meaning: A Sociosemiotics of Material Culture." In *Thinking Through Material Culture: An Interdisciplinary Perspective*. Philadelphia: University of Pennsylvania Press, 2005.
- Lanier, Jaron. *You are Not a Gadget: A Manifesto*. New York: Knopf, 2010.
- Law, John. "After ANT: Complexity, Naming, and Topology." In *Actor Network Theory and After*, edited by John Law and John Hassard. Oxford: Blackwell, 1999.
- . "Notes on the Theory of the Actor-Network: Ordering, Strategy and Heterogeneity." *Systems Practice* 5 (1992).
- Miles, Russ, and Dan Pilone. "The Next Iteration: If it ain't broke ... you still better fix it." In *Head First Software Development*. www.safaribooksonline.com/library/view/head-first-software/9780596527358/ch11.html.
- Nakamura, Lisa. *Digitizing Race*. Minneapolis: University of Minnesota Press, 2008.
- "Namespaces in XML 1.0 (Third Edition)." W3C. www.w3.org/TR/REC-xml-names/.
- Nowviskie, Bethany. "A Game Nonetheless." March 16, 2015. <http://nowviskie.org/2015/a-game-nonetheless/#more-2719>.
- "P5: Guidelines for Electronic Text Encoding and Interchange." The TEI Guidelines. www.tei-c.org/release/doc/tei-p5-doc/en/html/.
- Peirce, Charles Sanders. *Collected Papers of Charles Sanders Peirce*. Vol. 2. Edited by Charles Hawthorne and Paul Weiss. Cambridge, MA: Harvard University Press, 1960.
- . *The Essential Peirce*. Vol. 2. Edited by the Peirce Edition Project. Bloomington: Indiana University Press, 1998.
- . *Writings of Charles Sanders Peirce: A Chronological Edition*. Vol. 2. Edited by the Peirce Edition Project. Bloomington: Indiana University Press, 1984.
- Ramsay, Stephen, and Geoffrey Rockwell. "Developing Things: Notes toward an Epistemology of Building in the Digital Humanities." *Debates in the Digital Humanities*, 2012, 75–84.
- Radio Corporation of America, Advertisement, *Popular Science* (1950): 96.
- Saussure, Ferdinand de. *Cours de linguistique generale*. Edited by Tuillio de Mauro. Paris: Payot, 1986.
- . *Course in General Linguistics*. Translated by Roy Harris. London: Duckworth, 1983.
- Sponsler, Claire. *The Queen's Dumbshows: John Lydgate and the Making of Early Theater*. Philadelphia: University of Pennsylvania Press, 2014.

Text Encoding Initiative Consortium. "4. Default Text Structure—The TEI Guidelines." www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html.

———. "11.2.2 Embedded Transcription." www.tei-c.org/release/doc/tei-p5-doc/en/html/PH.html#PHZLAB.

———. "TEI element sourceDoc." www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-sourceDoc.html.

Winner, Langdon. "Do Artifacts Have Politics?" *Daedalus* 109 (1980): 121–36.

Chapter 6

ENCODING AND DECODING MACHAUT¹

TAMSYN MAHONEY-STEEL

Introduction

IT COULD BE said that editing a medieval text for any kind of modern edition is an endeavour doomed to failure. In the case of medieval philology and hermeneutics it can be a necessary and sometimes creative failure that may lead us to new insights. Indeed, it depends upon what we think we are doing in creating an edition, or even reading it, as to whether we believe that edition “fails.” This chapter explores the virtual divide between some of the manuscripts of fourteenth-century poet-composer Guillaume de Machaut, and a digital archive of late-medieval, French lyrics, *Je Chante Ung Chant*, which contains a number of Machaut’s fixed-form poems. I analyze the success of the archive as an edition or collection of poetry, and I look at the relationship between it and Machaut’s works and question how, and indeed if, we can compare archive and original source, and what criteria we can use for judgment. In doing so I will draw upon the work of Toril Moi, whose thoughts on the nature of authorial intent help to uncover an implicit dualism in editorial theory, and offer an alternative view of the relationship between edition and source material.

The Challenge of Editing Medieval Manuscripts

Medieval texts and codices were not authored with mass, uniform copying in mind. Indeed, they cannot be spoken of as having “authors” in the sense of the modern book, since the process of creation was idiosyncratic and influenced by the scribes, illuminators, patrons, and even the bookmakers (although *auctoritas* was nonetheless an important concept—see Timothy L. Stinson’s chapter in this volume). Each codex is a unique expression of its contents, as Stephen G. Nichols notes “the manuscript matrix is a place of radical contingencies.”² For the editor of medieval material, trying to rationalize what can be hundreds of unique and multimodal witnesses into one black-and-white page of text is a task as misleading as it is difficult. The *Roman de la Rose* Digital Library³ demonstrates

¹ I am indebted to the members of the Andrew W. Mellon Foundation Research Seminar participants at Johns Hopkins University, who gave me invaluable feedback on earlier versions of this chapter and helped shape my ideas. I owe particular thanks to Prof. Gabrielle Spiegel who has directed this seminar series over the past few years and enabled an atmosphere of collegial support for early career researchers.

² Nichols, “Introduction: Philology in a Manuscript Culture,” 9.

³ www.romandelarose.org

the complexity of the task: with around 130 digital surrogates of a possible near 300 manuscripts and early print editions, the *Rose* Library bears witness to over 300 years' history of copying, transmission, artistic style, and reception. The ability to scrutinize so many manuscripts via portals such as this has given rise to a new era in philological studies, allowing us to approach medieval texts by multimodal means, and to look beyond the confines of the black-and-white page. And yet, even with access to so many manuscripts, trying to understand a poem such as the *Rose* as *one thing* is to misunderstand how medieval writing and authorship was contemporaneously understood. We might as well seek to find the Platonic idea of the *Rose* as to recreate it in a single edition designed to capture all its nuances: we can say that there are as many *Roses* as there are manuscript witnesses to it.

We try to place as much information as we can into a single edition, intending to capture for our audience as many data as possible and attempting to convey, via this barrage of material, the full context of a medieval text. We include apparatus to show variants, we might include black-and-white reproductions of images, perhaps colour plates from one or two of the manuscripts, hoping that the reader can piece together these clues to extrapolate something about the contextual nature of the text. Some edition makers recognize the futility of the task and limit themselves to one version of a text (the approach championed by Joseph Bédier); some piece together what they believe to be an archetype text by reconciling variants (the Lachmannian method).⁴ The task is no easier with one manuscript, in fact the task is no easier even with only one *extant* manuscript, for the loss of information in the translation from parchment to page or screen, is still great.⁵ If, then, we are fated to endure this creative failure with each attempt, what then is it we are trying to do, what is it we are producing in the attempt, how does the end product relate to the manuscripts with which we are dealing (or, often, their digital surrogates), and does a digital edition provide us with any advantages? Before I try to answer these questions I will introduce the archive, and our example poet-composer, Guillaume de Machaut.

***Je Chante Ung Chant* and the Text Encoding Initiative**

Je Chante Ung Chant is a web-based archive of late-medieval French lyrics drawn largely from the fourteenth-century song and motet repertory.⁶ Its contents are gathered from *chansonniers* (miscellanies of notated *formes fixes* lyrics), complete works' manuscripts of composers such as Machaut, and collections of unnotated lyrics, which were eligible for musical setting or were cited in other notated lyrics, or whose forms suggest they

⁴ Bédier, "La tradition manuscrite," 161–196, 321–356; Timpanaro, *The Genesis of Lachmann's Method*.

⁵ See Timothy L. Stinson's chapter in this volume for a discussion of Bernard Cerquiglini's critique of Bédier and Lachmann, the issues of the variability of *variance* itself, and—drawing on Keith Busby—the problems of considering the instability of medieval works in the face of only one extant manuscript.

⁶ jechante.exeter.ac.uk/archive/

could have been notated. Created as part of an Arts and Humanities Research Council (AHRC) funded project by Gary Stringer, Giuliano Di Bacco, and Yolanda Plumley, the purpose of the archive is to explore citation and allusion in late-medieval French lyrics. The site features options to browse by first line (incipit), title, manuscript witnesses, and author, with further capabilities—such as browsing by textual or musical form—under development. The textual search function allows use of Boolean operators and Levenshtein distance (fuzziness), but does not currently facilitate the search of variant readings within texts, albeit many of these are encoded or in the process of being so.

Created with the objective of studying textual quotation, the archive naturally privileges those texts that were already being studied by the project investigators, although work continues to add to the corpus with the intent of producing an exhaustive set of extant French lyrics from ca. 1280–1420. This choice precludes texts that might cite across language borders, and bilingual or macaronic compositions.

Je Chante makes use of the Text Encoding Initiative (TEI) P5-conformant XML to mark up the lyrics and present an electronic edition of medieval poetry. The TEI consortium “collectively develops and maintains a standard for the representation of texts in digital form.”⁷ It uses XML as the encoding language to express the elements and attributes that the TEI defines in its guidelines, which are a set of recommendations for how the TEI might be used to express textual features in a document (the current iteration of the guidelines is known as P5). For the purposes of *Je Chante* the P5 schema has been only minimally modified with the intention of providing texts that are portable to other platforms and repurposable. The editorial policy, where more than one copy of a text is extant, is to choose a base manuscript—the one considered the most viable, reliable, or complete and to parallel encode variant witnesses within the base, as seen below.

For each text, three tabs with subsets of different views allow for a very comprehensive experience. The “Edition” tab with “Standard” view shows the text as it has been edited by the project team (see Figure 6.1); “Scholar” view makes available alternate spellings and readings, including those from older editions; “Metrical” view gives rhyme scheme and meter information. The “Manuscripts” tab allows a base manuscript and “Diplomatic” view as well as each alternate version from any other available witnesses. Finally, the “TEI P5” tab gives users the opportunity to view the encoding of the text—a view rarely offered by other electronic editions.

Figure 6.2 shows the encoding of stanza one of the ballade “Helas! pour quoy virent onques mi oueil” (henceforth “Lo53”) by Machaut, a lyric from his unnotated collection otherwise known as the *Loanges des Dames*. The base manuscript here is BnF fr 1586 (known as MS C—see Figure 6.3). Figure 6.4 shows the text and translation of Lo53, stanza one, without mark-up.

For each textual or spelling variant, apparatus (here denoted by the tag <app>) is applied to the word in question. The tag <lem> holds the reading from the chosen base manuscript, while the tag <rdg> holds the variant readings, with the tag’s attribute values (here @wit) supplying the short names of the MSS from which the variant readings

7 The Text Encoding Initiative Consortium, “TEI: Text Encoding Initiative.”

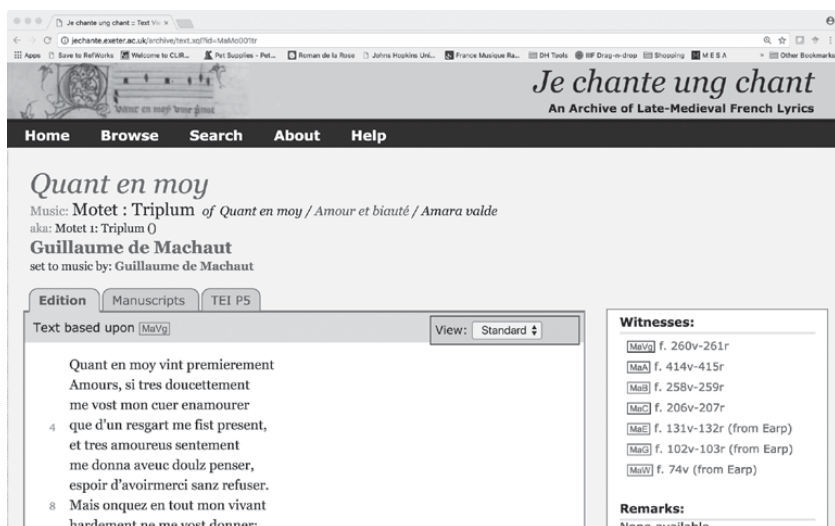


Figure 6.1. Screenshot of jechante.exeter.ac.uk showing the triplum text of the motet *Quant en moy / Amour et biauté / Amara valde* using the “Edition” tab and “Standard” view.

```
<lg type="stanza">
  <l n="1" met="10">Helas pour quoy vire<expan>n</expan>t o<expan>n</expan>q<expan>ue</expan>s
    <app><lem>mi</lem><rdg wit="#MSMaE">my</rdg></app>
    <app><lem>o<rhyme label="a">ueil</rhyme></lem><rdg wit="#MSMaB">euil</rdg><rdg wit="#MSMaE">
    >oeul</rdg></app></l>
  <l n="2" met="10"><app><lem>biaute</lem><rdg wit="#MSMaB">Beaute</rdg></app> p<expan>our</expan>
  moy decevoir <expan>et</expan>
    <app><lem>tra<rhyme label="b">ir</rhyme></lem><rdg wit="#MSMaB">trahir</rdg></app></l>
  <l n="3" met="10">ne <app><lem>cuer</lem><rdg wit="#MSMaVg #MSMaB #MSMaE">corps</rdg></app>
    <app><lem>cha<expan>n</expan>ge<expan>n</expan>t</lem><rdg
      wit="#MSMaVg #MSMaA #MSMaG #MSMaE">chantant</rdg></app> ne <app><lem>corps</lem><rdg
      wit="#MSMaVg #MSMaB #MSMaE">cuer</rdg></app>
    <app><lem>plei<expan>n</expan></lem><rdg wit="#MSMaE">plain</rdg></app> de tel v<rhyme
      label="a">ueil</rhyme></l>
  <l n="4" met="10">Qui <app><lem>fei<expan>n</expan>t</lem><rdg wit="#MSMaA #MSMaG #MSMaE">
    >faint</rdg></app> amer <expan>et</expan> ne fait q<expan>ue</expan>
    <app><lem>ha<rhyme label="b">ir</rhyme></lem><rdg wit="#MSMaB">air</rdg></app></l>
  <l n="5" met="10"><app><lem>Miex</lem><rdg wit="#MSMaB">Mieulx</rdg></app> me
    <app><lem>vausist</lem><rdg wit="#MSMaB #MSMaE">vausist</rdg></app> estre
    <app><lem>nez</lem><rdg wit="#MSMaG">nes</rdg></app>
    <app><lem>sa<expan>n</expan>s</lem><rdg wit="#MSMaB">sanz</rdg></app> ment<rhyme label="b">
    >ir</rhyme></l>
  <l n="6" met="10">Sa<expan>n</expan>s <app><lem>yex</lem><rdg wit="#MSMaB">ieulx</rdg><rdg
    wit="#MSMaG">iex</rdg><rdg wit="#MSMaE">yeux</rdg></app>
    q<expan>u<reg></reg>a</expan>mer dame ou ta<expan>n</expan>t <app><lem>truis</lem><rdg
    wit="#MSMaE">truis</rdg></app>
    <expan>contr</expan><rhyme label="c">a</expan>ire</rhyme></l>
  <lg type="refrain">
    <l n="7" met="10">Q<expan>ua</expan>nt <app><lem>loyaute</lem><rdg wit="#MSMaB">
    >loyaute</rdg><rdg wit="#MSMaE">loyaltes</rdg></app> ne
    <app><lem>mai<expan>n</expan>t</lem><rdg wit="#MSMaA #MSMaG">
    >meint</rdg></app> en son vi<rhyme label="c">aire</rhyme></l>
  </lg>
</lg>
```

Figure 6.2. TEI P5 conformant XML encoding of stanza one of “Helas! pour quoy virent onques mi oueil” (Lo53).

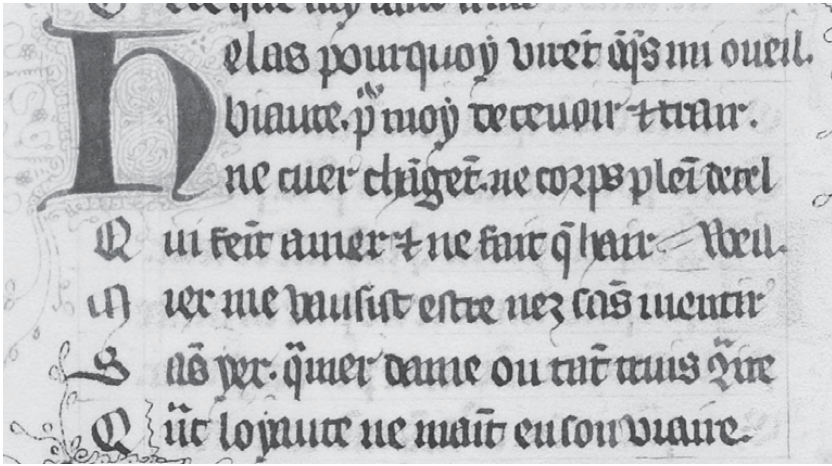


Figure 6.3. Stanza one of “Helas! pour quoy virent onques mi oueil” (Lo53) as it appears in MS C. © Bibliotheque nationale de France. Reproduced by permission.

Helas pour quoy virent onques mi oueil
 Biaute pour moy decevoir et trair
 Ne cuer changent ne corps plein de tel vueil
 Qui feint amer et ne fait que hair
 Miex me vausist estre nez sans mentir
 Sans yex qu’amer dame ou tant truis contraire
 Quant loyaute ne maint en son viaire

[Alas! Why did my eyes, in order to deceive / and betray me, ever behold beauty? / I wish for neither an inconstant body, nor such a heart, / that feigns love and only creates hate. / Without a lie it would be better to have been born / without eyes, than to love a lady in whom I find such contrariness, / since loyalty does not reside in her countenance.]⁸

Figure 6.4. Text and translation of stanza one “Helas! pour quoy virent onques mi oueil” (Lo53) without mark-up.⁹

are taken (this method is known as parallel segmentation). Thus, when we choose to view the version in MS BnF fr 1584 (MS A), the *Je Chante* platform takes the text as it is found in MS C, and replaces every <lem> coded word with the <rdg> assigned to MS A (if there is no <rdg> for MS A, then it is assumed that the <lem> is the same). It is an act

⁸ My translation.

⁹ My translation.

of substitution of individual words, and not one of drawing on a separate version of the text. This method makes it easy to articulate individual differences between texts, but does not treat alternate versions of a text as discrete entities.

The Challenges of Digital Editing

As with any edition making, encoding a text using the TEI can be fraught with challenges. On the other hand, the advantage of digital editions or collections is that they may have the luxury (or constraint) of having a specific purpose. As Julia Flanders notes “[t]he space of the digital archive is scoped and bounded by the editorial problem it is designed to address.”¹⁰ In the case of *Je Chante* it was “devised as a research tool for the study of citational practices in the French lyric repertory from ca. 1280–1420.”¹¹ Scholarly editing, according to Timothy L. Stinson, “usually entails many acts, or at least claims, of supersession.”¹² That is, a new edition is frequently created because something is perceived as lacking in previous editions (if they exist), or even in the original source. This is particularly true of print editions where the aim is to create something that conveys the most information in the most comprehensive yet compact manner. Economic reasons are usually the deciding factor. A scholarly publication must be both palatable and affordable to the broadest possible audience (researchers, students, and other interested parties), while maintaining rigorous editorial standards that will stand up to peer scrutiny. Digital editions often have fewer economic constraints since they are frequently grant funded and made freely available rather than sold like a book; this can mean that they have the power to answer demands that are less global. The *Je Chante* archive, for example, is not an attempt to supersede other editions of the lyrics it contains, but rather to bring a wide range of lyrics together under one system to enable search across a number of lyric repertories. Indeed, without the music of the notated lyrics, the archive is arguably offering something less comprehensive than editions containing the scores.

Code is not neutral.¹³ When we make a choice as to the elements and attributes we use to define a text, we automatically limit alternate possibilities for that definition. For example, the encoding of variants in the *Je Chante* archive was carried out by means of parallel segmentation. As explained above, this results in one text, with alternate versions encoded within that text rather than having the text of each witness encoded separately. So, while we can view the different versions of a particular text via the program that renders them, only one text is in fact being drawn upon, whereas the actual process of encoding drew upon several witnesses.

¹⁰ Flanders, “Rethinking Collections,” 168.

¹¹ *Je Chante Ung Chant*, “About.”

¹² Stinson, “Makeres of the Mind,” 39.

¹³ See, for example, Ted Underwood’s discussion of the bias inherent in algorithms and their use: Underwood, “Theorizing Research Practices,” 64–72.

The choice we made to encode variants using the parallel segmentation method is an interesting one. Why not encode each text separately, thus respecting the individuality of the witness? Each version of a lyric, however minimally different, could have been encoded as a separate text, but this could have wildly skewed the search results. A method would have had to have been found to eliminate from the search process copies of the same text as it appears in different MSS. It would simply not have made sense to treat different MS versions as different texts when the purpose is to explore citational links between lyrics. And, in eliminating copies, we would have had to have chosen a base MS, which brings us full circle to what we had already chosen to do. But, if we are only searching the base MS text, why are we encoding variants at all? This is a far more interesting question. One aim of the project was to create a set of texts that are “portable,” i.e. they can potentially be reused in other archives with different purposes.¹⁴ In this respect including variants was one way to maximize the amount of information conveyed by our digital edition. Why then not encode scribal hand, rubrics, marginalia, links to PURLS (Personalized Uniform Resource Locator) of images, etc.? The choice may well point to our assumptions about what is “normal” in an edition: variants are the most usual inhabitants of any apparatus in scholarly hard copy editions. As Julia Flanders asks: “how does the digital collection inherit or re-echo constraints that its virtual status should be able to transcend?”¹⁵ Our desire to encode variants, and also not to regularize spellings (which would have aided discovery of citation), seems to be such a constraint that could be transcended. It could also be argued that our choice made the texts less portable because they come with a pre-chosen base text and variants encoded in the poetic line. Anyone repurposing the texts who did not desire this manner of conveying apparatus would need to perform some fairly complex XSLT transformations to separate out one parallel-encoded file into its constituent sources.

To illustrate further the problems of encoding, let me introduce a case study from my own teaching. In January 2016 I taught a short introductory course to the Text Encoding Initiative. As part of the learning experience the students were given archival material to encode. The material was the “exclusion files” from the Roland Park Company papers. Located in Baltimore, the Roland Park Company was responsible for the first specially designed suburban community in the United States.¹⁶ The exclusion files are a set of forms detailing the interactions of plot salesmen with prospective clients that resulted in the latter being refused the chance to purchase (see [Figure 6.5](#)). The refusal was usually on grounds of race and class, and as such these documents give a fascinating, if sobering, insight into the attitudes of a particular place and era.

The form lays out the information in what might be called a typical manner. The prospective client’s name followed by their contact details for home and work. The form then gives space to detail the kind of property the client desires, followed by the names of the salesmen and brokers involved. Below the section that requires specific data is a space for

14 *Je Chante Ung Chant*, “Encoding.”

15 Flanders, “Rethinking Collections,” 165.

16 Mowbray and Bouton, Roland Park Company Records. My thanks go to archivist Valerie Addonizio for introducing my students to these materials.

FORM CITY 108 1900 4-11-30 M. & M. CO.

NON-OPERATOR *Exclusion file*

Prospect	[REDACTED]	Listed	1/26/31	19	No.	2416
Home Address	[REDACTED]	Tel. Lib.	5855			
Position	Cashier	Firm	Equit. Life Assur. Society			
Business Address	Equitable Bldg.		Tel. Pl. 6214			
Requirements:	3 BR, 1 BA					
Salesman 1.	[REDACTED]	2.	3.			
Broker						
1/26/31	They came to RS-1. Seemed very much interested and said they were not living in an apartment and might consider change. EWD					
2/20	Phoned. Said their lease expires in October but are not sure whether they will buy a house at this time. EWD					
5/21	Visit. Called upon [REDACTED] and find she is a Hebrew. EWD					

Figure 6.5. Example Exclusion File from the Roland Park Company Papers, Special Collections and Archives, Sheridan Libraries, Johns Hopkins University.

a narrative of interactions. How might one encode this? For the purposes of our class, we wanted to capture the content rather than the appearance of the document, which meant we could dispense with some of the physically descriptive tags and attributes. Other than the order of the data on the form, there is no explicit hierarchy to the document. So, one could simply tag each piece of information in order to render it searchable. Figure 6.6 gives a simplified idea of how this might work, with the elements <person>, <date>, <residence>, and <num> having equal standing within the <front> element.¹⁷

However, my students chose to look for the implicit hierarchy in the document. They saw the form as a set of interactions between two or more human beings. Therefore, they opted to encode the top section of the documents according to the people recorded (the prospect, the salesmen, the broker), with any information associated with that individual as child elements of the person element (see Figure 6.7).

Encoding is not a one to one translation. As this example from my class shows, not only are there different choices that can be made with a system such as the TEI, these choices are frequently based on an act of interpretation. It can also be representative of a particular historical moment or political mind-set. My students were encoding these sensitive documents less than a year after the protests in Baltimore over the death in police custody of Freddie Gray.¹⁸ Furthermore, Johns Hopkins University, as part of its *Ten by Twenty* campaign goals and HopkinsLocal initiative,¹⁹ is committed to finding means to work more closely with its local community and local people. It would not

¹⁷ We created the element <caseFile> for this project as the TEI did not have anything approximating this kind of document in its guidelines.

¹⁸ "Timeline: Freddie Gray's Arrest, Death and the Aftermath," *Baltimore Sun*.

¹⁹ "Commitment to our Communities," *Johns Hopkins University Progress Report 2015*; Johns Hopkins University and Health System, "HopkinsLocal."

```

<caseFile>
  <front>
    <person role="prospectiveClient">
      <persName>J. Doe</persName>
    </person>
    <date when="1930-01-01">January 1, 1930</date>
    <residence>
      <address>
        <addrLine>1 Main Street</addrLine>
      </address>
    </residence>
    <num type="phone">12345</num>
    ETC...
  </front>
</caseFile>

```

Figure 6.6. Encoding with elements each having equal hierarchy.

```

<caseFile>
  <front>
    <person role="prospectiveClient">
      <persName>J. Doe</persName>
      <date when="1930-01-01">January 1, 1930</date>
      <residence>
        <address>
          <addrLine>1 Main Street</addrLine>
        </address>
        <num type="phone">12345</num>
      </residence>
      ETC...
    </person>
    <person role="salesman">
      <persName>J. Smith</persName>
    </person>
    ETC...
  </front>
</caseFile>

```

Figure 6.7. Encoding with details nested under <person>.

be surprising if these considerations had, consciously or unconsciously, influenced the students' desire to make their encoding of the Roland Park Company Exclusion Files more human-centric. It is important to bear all this in mind, for even coding something as seemingly hermeneutically unproblematic as a line or a line group in poetry will have implications for the way in which the document is searchable, or for what other features

can be encoded, and may be reflective of contemporary experiences in the lives of the editors, however neutral that may purport to be.

Guillaume de Machaut

Much has been written on the authorial persona of fourteenth-century poet and composer Guillaume de Machaut.²⁰ It is not my intention here to delve into the extent to which Machaut's presence within his own texts and manuscripts can be traced and analyzed, however, it is important to remark upon because the perception of his authorial intention is a factor influencing how editors might approach his works. For example, the forthcoming complete works' edition of Machaut's oeuvre (in print rather than digital format) is using MS A as its base manuscript because this is believed to be the most authoritative. It contains one of the most complete versions of his opus, has an iconographical program that may imply authorial overview,²¹ and contains a tantalizing rubric (discussed below). It is also one of only a handful of manuscripts to contain his *Prologue*; in which Machaut, retroactively engineering a fictionalized *raison d'être* for his entire oeuvre, imagines an encounter with Nature and her children and posits himself as the recipient of her orders to write poetry and music. Nature addresses Machaut by means of a *ballade*, and entrusts her children (Meaning, Rhetoric, and Music) to him in the first stanza:

Je, Nature, par qui tout est fourmé,
 Quanque a ça jus et sur terre et en mer,
 Vieng ci a toy, Guillem, qui fourmé
 T'ai a part pour faire par toi fourmer
 Nouveaux dis amoureux plaisans.
 Pour ce te bail ci trois de mes enfans
 Qui t'en donront la pratique,
 Et, se tu n'es d'eulz trois bien cognoissans,
 Nommé sont Sens, Retorique, et Musique.

[I, Nature, by whom all things took form, / Whatever is above and on the earth, in the sea, / Come here to you, Guillaume, a man / I created especially to create / New and pleasant poems about love. / And so I entrust you with three of my children / Who will give you practical knowledge, / And, if you do not recognize them readily, / Their names are Meaning, Rhetoric, and Music.]²²

Written near the end of his life and placed as a splendid opening to four of the extant manuscripts devoted to his works the *Prologue* frames Machaut's works, career, and authorial persona. It consists of four *ballades*—Machaut being addressed by and replying to Nature, and then being addressed by and replying to Love—followed by a

²⁰ In particular, see Brownlee, *Poetic Identity* and Palmer, "The Metafictional Machaut," 23–39.

²¹ Leo, "Illuminating the *Voir Dit*."

²² Machaut, *The Fountain of Love*, trans. Palmer, 3.

longer poem just short of 200 lines that meditates on his commission. In addition to providing a schema for his poetical and musical output, the *Prologue* also implies how his works should be transmitted and received, emphasizing the holistic and thematic qualities of his corpus: these works *belong* together and they have certain elements in common. In providing this conceit for his complete works' collections, Machaut is at once showing himself to be a humble vector for the professed will of the allegorical figures of Nature and Love, and for the purpose of praising ladies, while also subtly aggrandizing his position as an author-figure in command of his output.

Yet each of the four manuscripts that contain the *Prologue* witnesses a different version of Machaut's oeuvre. None of these complete works' manuscripts actually contain his entire output; there are variations—albeit fairly minor—between them. Each manuscript has its own constellation of reasons and factors as to why it exists. Put in the words of Moi, each manuscript is an action and expression that was made in a particular time and place, by a person (in the case of manuscripts a group of persons), and which has its own effect on the world.²³ In recognizing the unique nature of each MS, I am not fetishizing them as objects, rather I am acknowledging each one as an individual expression of a set of circumstances.

Let us take for example Machaut MS BnF fr 1584 (MS A), whose index tells us: "Vesci, l'ordenance que G. De Machaut vuet qu'il ont en son livre" —"Here is the order which G. De Machaut wants there to be in his book." This tantalizing clue, citing both his name and emphasizing it with a personal pronoun, has led scholars to surmise that he oversaw the production of MS A.²⁴ Indeed, MS A was produced in his lifetime and the sumptuous illustrations accompanying the *Prologue* are believed to contain a real portrait of the poet.²⁵ The rubric is followed on the line below by the word "Premiers" [first]. Is this indicating which work comes first? Or is it an adjective to "livre" implying that it is only the first book (i.e. an original subsection of this manuscript) to which the ordering pertains? Or is this Machaut's first book that he has produced?

The last option seems unlikely since MS C was produced much earlier in his lifetime, unless the reference is to Machaut's first complete works' book—although there is no explicit indication that Machaut understood the notion of "complete works" in quite the way that we do.²⁶ If the rubric implies a first book, as in a section of the MS, then there

²³ Moi, *Revolution of the Ordinary*.

²⁴ Williams, "An Author's Role," 66–67.

²⁵ The likelihood of this being an actual author portrait seems strong. Machaut is depicted in one of the images with amblyopia, and in the *Voir Dit* describes himself as Toute-Belle's "one-eyed servant." [vostre borgne valet] (Machaut, *Voir Dit*, 206–7).

²⁶ MS BnF fr 1584 (MS A) certainly does contain the majority of Machaut's works, and the choreography of the manuscript's layout, the intratextual references, and the performance of his authorial persona have led him to be described as someone who foreshadows modern authorship (see, for example Palmer's essay, "Guillaume de Machaut and the Classical Tradition," 241–260). In the *Voir Dit*, he describes a book that "contains all the works I have ever composed" [*ou toutes les choses sont que ie fis*] (Machaut, *Voir Dit*, 124–125), which is in the process of being prepared for one of his patrons.

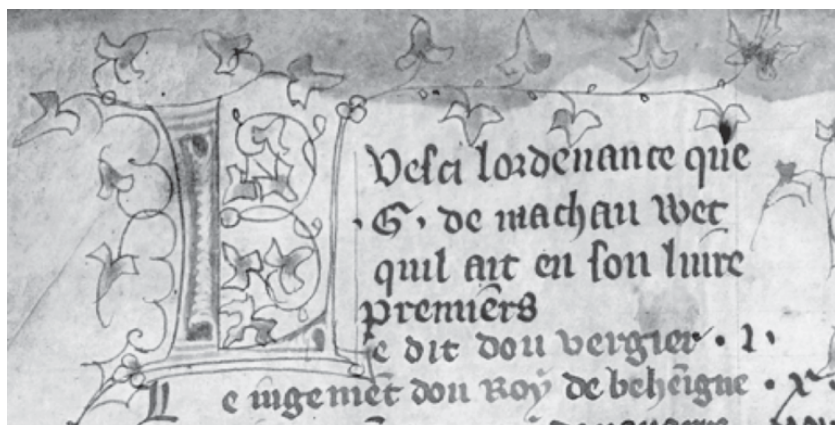


Figure 6.8. Rubric from the index of MS BnF fr 1584, f. Av. © Bibliothèque nationale de France. Reproduced by permission.

is no further indication as to the scope of this “book” or how many make up the MS. “Premiers” could well denote the starting point, although being in red and immediately following the rubric, it does cause some confusion. If we look further at the rubric, we might question what is meant by “ordenance.” Does it imply that this organization of every song, poem, and dit is precisely what Machaut wanted (at least at that moment) or was it a more general sense of order? In several of his manuscripts a set of narrative dits are followed by the unnotated lyrics, which in turn are followed by the music. Even when this pattern isn’t followed, the music, narrative dits, and unnotated lyrics respectively tend to be collected together. Perhaps “ordenance” then can be read more approximately as genre-based collation.

Aside from the paratextual clues of the rubrication, Machaut appears to insinuate himself into and attempt to disrupt the reading practices of his audience. In *Controlling Readers: Guillaume de Machaut and His Late Medieval Audience* Deborah McGrady explores the complex relational matrix between reader, author, and book producers (makers, scribes, etc.). Specifically, taking her cue from Robert Jauss, she emphasizes how readers are a force that shapes textual meaning, book creation, and imagery. McGrady suggests that we read the cues in text and image and book production that tell us how the creators of those things saw their readers. She argues that “Machaut’s self-conscious articulation of authorship emerged in response to a perceived aggressive audience that threatened to appropriate and rewrite his text at every turn.”²⁷

In particular, McGrady examines Machaut’s *Livre dou Voir Dit* (Book of the True Tale) as a space in which public and private reading practices collide and in which Machaut examines the problems of his own agency and control over his work. The *Voir Dit* is a purportedly true tale of the love affair between the aging Machaut and a young admirer whom he calls “Toute-Belle” (All-Beautiful). It combines poetic narrative, letters exchanged

²⁷ McGrady, *Controlling Readers*, 9.

between the couple, and short fixed form poetry that they write for one another (there is also reference to musical settings of some of the poems, but the notation is not included in the narrative). What ostensibly begins as a typical medieval love story is in fact a metaphor for Machaut's relationship with his supposed interfering audience. McGrady contends that Tote-Belle, rather than being the subservient student to Machaut's genius that she purports to be, is in fact an ambitious and intervening editor and co-author:

Tote-Belle's careful reading and subsequent adroit imitation of the poet are compared to the medieval practice of scraping and erasing previous writings on parchment to inscribe new compositions ... In essence, Tote-Belle's mastery of the skills taught by Guillaume threatens to render the master irrelevant not only to his student, but to poetry in general and specifically to the *Voir Dit*.²⁸

Machaut's strategy for reasserting his authority is to create a narrative that weaves together different modes of interacting with the text thus disrupting the experience of it by being neither fully for listening or reading:

The sixty-three poems and ten songs invite an oral performance, while deictic references to the book's materiality, forty-six private prose letters, and elaborate pictorial programs privilege a physical encounter with the book.²⁹

Furthermore, he deliberately takes the reader outside the *Voir Dit* text on a journey that still keeps the reader within the confines of his oeuvre.

Des autres choses vous diray
Se diligemment les querez
Sans faillir vous les trouverez
Aveques les choses notees
Et les balades non chantees

[Concerning the other compositions, I tell you / That to find them without fail
/ You should diligently look / Among the poems set to music / And the ballads
not meant to be sung]³⁰

None of the poems in the *Voir Dit* mentioned as being set to music are in fact notated in the body of the narrative. One must turn to a different section of the codex entirely to find these, and yet another to find the others referred to as "not meant to be sung." Machaut is clearly envisaging a situation in which a reader has access to a manuscript that contains not only the *Voir Dit*, but his musical works and other poems. The further implication is that the *Voir Dit* loses some of its sense if it is transmitted without these.

In addition to his more overt clues to the reader to encourage engagement with and journey through the physical repository of his texts, Machaut further instigates scrutiny

²⁸ McGrady, *Controlling Readers*, 54.

²⁹ McGrady, *Controlling Readers*, 8.

³⁰ Machaut, *Le Livre dou Voir Dit*, trans. Palmer, ll. 521–525.

of his works by his reuse of material throughout his oeuvre. This self-citation may prompt the canny reader to engage both intellectually and physically with Machaut's works in a contrastive mode to anticipated linear reading practices. Plumley ponders whether these intratextualities that infuse his corpus would have been recognized by the reader, and whether Machaut intended them to be read in this hypertextual manner. Her in-depth analysis of Machaut's *Loanges* lyrics, with which I am in agreement, shows that he was "seeking to orchestrate his audience's responses by creating an intratextual dialectic that invited the cross-reading of individual texts."³¹ Machaut uses sets of adjacent lyrics constituted from similar themes and vocabulary (for example Lo71, *Puisse que Desirs ne me laisse durer*, and Lo72, *Puisse que Desirs me veut dou tout grever*), and remotely placed lyrics (such as Lo71 and Lo140, *Biauté parfaite et bonté souverainne*, the latter of which uses the former's incipit as its refrain) to create programs of interconnection within his body of work.³² While there are sets of juxtaposed lyrics that share thematic material, Plumley notes that the overall organization of the *Loanges* lyrics is largely heterogeneous and likely approximately chronologically ordered.³³ This vindicates the idea that Machaut desired his audience to engage in an explorative manner with his lyrics, for they were not organized according to theme and citation, but rather left in tantalizing miscellaneous design to reward the discerning and adventurous reader.

Machaut, then, emerges from the manuscripts of his works as an author with a design for the presentation and transmission of his work. But, is this something we can or indeed should capture in any edition of his corpus? What possibilities does the digital realm afford us? And, how do we understand the relationship between a controlling medieval author, and modern digital renditions of his poetry in an online archive? To answer some of these questions we need first to delineate what is understood by the intentionality of an author or the matrix of intentionality generated by those who bring a manuscript into being.

Intentionality and Editorial Theory

Latent within our theories of editing and efforts as editors is an unnecessary dualism. Just as philosophers have posited and tried to overcome mind-body and mind-world dualism, so we can posit and attempt to surmount such a rift between the edition and its source. Whether or not we subscribe to Bédier's idea of the 'best text', or Karl Lachmann's stemmatics, the text, in some fashion or another, hovers at our shoulder like a ghostly entity, demanding to be perfectly rendered and expressed in our edition. We console ourselves that a one-to-one translation from source to edition is never possible, and that somehow with each attempt at editing we must try to come closer to the source, even though like Achilles we can never pass the tortoise. Trying to express ideally a text in an edition is as flawed as trying to express authorial intention as if it is something

31 Plumley, *The Art of Grafted Song*, 321.

32 Plumley, *The Art of Grafted Song*, 322–23.

33 Plumley, *The Art of Grafted Song*, 338–40.

pre-determined that is poured into the text-as-vessel. For Moi this view of intention as the expression of the pure will of the author distilled into text is as faulty as the New Critics' attempts to discredit it. Moi notes that the New Critics dispense with authorial intention and favour the reaction of the reader in understanding texts, but do not offer an alternative theory of intention, rather they cling to it as a taboo:

The New Critics postulate this picture of intentions in order to denounce it. They don't deny that authors *have* intentions, but claim that there is no way to establish what they might have been, and that in any case, readers have no business asking after them. The meaning of the text resides in the text itself, not in whatever intentions the author might once have had.³⁴

Yet texts and intention do not work in this way. The text is not a vessel for the will of the author, perfectly worked out in advance. As Moi rightly argues, the New Critics are only besting a straw man in offering this view of intentionality. For example, a novel is not pre-formed in our minds and then extracted in an alchemy of textualization from mind to page. "Writing is thinking. Texts grow and change as the writer works on her sentences."³⁵ The text expresses what has been written at a particular time and place, so it is an expression of those intentions, but it is not an object that encapsulates the will of an author as if this will were also an object against which we could measure our reactions. The text is an action that has an effect upon us. Put differently, asking questions of the text—i.e. why something was expressed in a certain way—is not something that takes us "outside" the text to commune with the ghost-like emanation of the authorial will, but something that leads us to a deeper engagement with the text itself, that leads us to be affected by it.³⁶

If we view editing as a special class of interpreting texts and intentions, and understand intention in Moi's manner, then we can free ourselves from some of the more intractable problems of edition making. Moi's approach extricates us from the author-text intentionality dualism, and leaves us with a text that itself has an immanent intentionality that acts upon us. In the case of a manuscript, this is a matrix of intentionality that springs forth from the combined actions of author, scribe, bookmaker, patron, and anyone else involved in the process. We might, therefore, think of the intentionality of the manuscript itself as a culmination of these respective inputs.

I discussed above the idea that medieval texts and their manuscripts are not created with mass identical copying in mind. They are, however, created with copying in mind. Transmission is part of the intentionality of the manuscript matrix. Machaut was eager to imply the unity of his output through strategies that lead the reader to other parts of his codex because he realized that interfering readers, scribes, and patrons would likely divide, repurpose, and even rewrite as they saw fit, which they

³⁴ Moi, *Revolution of the Ordinary*, 200.

³⁵ Moi, *Revolution of the Ordinary*, 201.

³⁶ Moi, "Reading as a Practice of Acknowledgment: The Text as Action and Expression," in *Revolution of the Ordinary*.

indeed did. We could argue that there is an intentionality inherent in the nature of the manuscript that is at odds with the intentionality expressed in Machaut's texts. How, then, do we respond to this as editors? To which intentionality, if any, are we beholden?

Editing is an action carried out in reaction to a text because it has had an action upon us. Moi describes texts as "actions and expressions," things that people have done, and argues therefore that texts place claims upon us.³⁷ The claim is one of acknowledgement, a term she draws from Stanley Cavell. Acknowledgement goes beyond recognition since it involves "that I *do* something or reveal something."³⁸ An edition, therefore, is an extreme acknowledgement of a text acting upon us. We are reacting to the communicative nature of a text by transmitting it, and, in the case of a manuscript, we are reacting to the expectation that the material therein be transmitted in a new way.

Machaut Encoded

The relationship between the *Je Chante* archive and the works and manuscripts of Machaut is a complex one. For Machaut's oeuvre, there is an implicit tension between the intentionality of the manuscripts, and that expressed overtly and tacitly in his texts. Manuscripts, as noted above, are unique objects that are created with the understanding that they will likely be copied—whole or in part. It is a factor of the intentionality built into the manuscript matrix that texts be transmitted in new ways. Machaut, as McGrady proposes, was attempting to resist a culture of scribes and readers who would happily eviscerate and repurpose texts. Ironically, his very vehicle for transmission was working against him: manuscripts have stronger intentions than authors. McGrady's recent project, *Machaut in the Book*, explores the many manuscripts containing his work in fragmentary formats, often unattributed. The archive acknowledges that tradition by placing Machaut in a new electronic *chansonnier* alongside contemporaries with whom he may not have previously shared folio or codex. Hence, the *Je Chante* archive acknowledges the manuscript intentionality by placing texts in a new environment, to be experienced in a new way; but it also acknowledges Machaut's tacit intentions by supplying a platform upon which intertextuality can be explored.

Digital environments can respond to and acknowledge multiple manuscripts by showing us the *variance* and *mouvance* between them in a manner not possible in a print edition. However, platforms like the *Je Chante* archive, which display variants alongside the base text and allow users to swap views, do not replicate the reading experience of a single manuscript in which we do not have access to variants—our appreciation of the text is in that case based exclusively on one witness, which would have been the most usual experience for the medieval reader.³⁹ Furthermore, while we can use the *Je Chante*

³⁷ Moi, "Reading as a Practice of Acknowledgment: The Text as Action and Expression," in *Revolution of the Ordinary*.

³⁸ Cavell, "Knowing and Acknowledging," 47–71, 257.

³⁹ See Stinson's chapter in this volume.

archive to trace potential lines of connection between texts, this is not the experience of intertextuality that the contemporary reader would have had; their experience would have been more embedded and intuitive. We are trying to recreate the experience of who knew what and when, but the archive cannot replace the experience of close reading the texts and knowing their manuscript sources. The archive can supply us with a set of commonalities between texts, but it cannot give us the rich and deep context that comes from the texts, their manuscripts, and the paratextual data that accompanies them. So while the archive acknowledges the intertextual culture that inheres in medieval lyrics and their manuscripts, it cannot provide a shortcut to understanding what citational culture means experientially.

While the archive cannot provide a cultural immersion (indeed no platform or edition could do this), it does reflect the ludic nature of the citational culture with which it engages. Both the archive and Machaut's output/manuscripts anticipate a certain level of knowledge from the user. The archive takes an aspect of medieval literary culture, namely citation and allusion, and uses it as a roadmap for discovery. As such it is responding to, and acknowledging the implicit game within those texts: a literary hide and seek. In turn, it asks us to respond by continuing to play the game, to use the search function to find the webs of interconnections within late fourteenth-century lyrics. As we have seen, this citational game is one that Machaut himself played avidly. Hence, while the archive does not respond to Machaut's attempts to have us transmit his corpus wholesale, it does acknowledge, in Cavell's sense, his action upon us in asking us to trace lines of reference between different works. The archive, in turn, places demands upon its user. As with Machaut's corpus, there is little in the way of instructions for use. A certain level of user knowledge is anticipated: it is not an introduction to medieval French lyrics; there are no translations; there are no hyperlinks to MS images; and beyond a short explanation of the search functionality, there is an assumption that the user is familiar with this, too.

In working with the archive, we must also appreciate it as a work in and of itself and not merely as a response or handmaiden to the manuscript sources. It has its own intentionality and influences, and these things, like Machaut's intentions and the impetus of manuscript culture, are sometimes in competition. The professed intention of the archive was for exploring citation and allusion, and this can be seen as clashing somewhat with our desire for the texts to be portable, which was interpreted by the team as meaning the texts should convey as much information as possible for other potential users of the data. That "portable" meant this can be ascribed to the project's being AHRC funded, since this manner of government funding requires that project data and outcomes be made available. In this regard, as editors, we wished to produce something comprehensive that could be used in a variety of ways. Like my students' encoding work, the manner of creating the archive was subject to a number of contingent and interpretative factors.

A digital edition can be more holistic than a print edition, containing all variants, manuscript images, translations, in fact anything relevant we can dream up and have the time to append. Yet, as we have seen with the *Je Chante* archive, a digital edition can also answer a specific question or problem. As such, its acknowledgement of its source

can be specific. Such specificity may not align with all the kinds of intention we can see in the source, but as I have noted, an edition cannot fully emulate a source and to try to do so constitutes a form of failure. Even a holistic edition is not capturing its sources completely—the act of moving the data to a new realm changes the context and thus our reaction to and acknowledgement of it. The *Je Chante* archive allows a user to assume a form of scribal-authorial role. Each search produces a sub-*chansonnier*, a collection of lyrics based around the search terms. At the moment, users are not offered the facility to save those searches or the lists produced, but this is worth considering for future iterations, enabling the curious user to create sets of digital editions for their own particular research needs.

The *Je Chante* archive creates something new that in turn requires an acknowledgement from us. In acknowledging the works of Machaut and the other poet-composers represented within it, it does not merely emulate. It cannot. What it does do is provide us with an alternative form of interaction with some texts, and helps to liberate us from some of the constraints of print editions such as what Julia Flanders calls the “artificial dichotomy between documentary and critical editorial approaches.”⁴⁰ Like many medieval manuscripts, the archive is a manufactured collection, and—despite the possibility of simultaneously viewing it from many locations—a unique one (albeit there is a backup server). In this regard, it is an instantiation in the manuscript tradition, a product of many hands, assembling a collection, with rules dictated by a patron (namely the AHRC, and, to some extent, the University of Exeter). It also provides us with a new kind of art form that we as users must acknowledge: each TEI-encoded lyric becomes a poem that praises the variant and the metadata of the poem, for all are captured together in one text. Given the uniqueness of each medieval manuscript, putting the variants together shows the intractable problem of editing the medieval manuscript, and the encoded poem, unrendered as we see it in the TEI P5 tab, is itself a poem to that problem. Even when we only have spelling variants, it shows that we cannot stack up the poems that constitute Lo53 neatly on top of one another and have them all line up. Indeed, for poems that have no variants there are still headings, genre-related information, corrections, regularizations etc. that form part of the body text.

Conclusion

Like a medieval manuscript, the *Je Chante* archive is a creation with which we can have a multifaceted relationship. Both have a form of intentionality that transcends any one author and that is representative of a complex set of interactions between multiple agents (author, scribe, illuminator, bookmaker for the manuscript, and programmer, encoder, editor, scholar for the archive). Machaut demands to be acknowledged through the holism of his work, the manuscripts containing his corpus demand to transmit their works through reading and through copying, the archive demands that we explore the interconnections between lyrics, to create our own sub-sets of them, but to also

⁴⁰ Flanders, “Rethinking Collections,” 168.

appreciate the intractable problems of editing medieval manuscripts. Together these demands sometimes harmonize, and sometimes clash—multiple layers that work together like one of Machaut's motets (polyphonic, polytextual pieces of music). If there is a conclusion to be had from all this, it is that emulation and translation are not possible. We are not dealing with objects that can be perfectly copied and reproduced. Rather, in Moi's words, we are dealing with actions and expressions that have effects upon us and the world. The answer is not to seek one perfect medium but to listen to the intentionality of the poet, the manuscript, and the digital archive in a non-static, harmonious relationship.

Bibliography

- Baltimore Sun*. "Timeline: Freddie Gray's Arrest, Death and the Aftermath." <http://data.baltimoresun.com/news/freddie-gray/>.
- Bédier, Joseph. "La tradition manuscrite du *Lai du l'Ombre*: Réflexions sur l'art d'éditer les anciens textes." *Romania* 54 (1928): 161–96, 321–56.
- Brownlee, Kevin. *Poetic Identity in Guillaume de Machaut*. Madison: University of Wisconsin Press, 1984.
- Cavell, Stanley. "Knowing and Acknowledging." In *The Cavell Reader*, edited by Stephen Mulhall. Cambridge, MA: Blackwell, 1996.
- Flanders, Julia. "Rethinking Collections." In *Advancing Digital Humanities: Research, Methods, Theories*, edited by Paul Longley Arthur and Katherine Bode, 163–74. New York: Palgrave Macmillan, 2014.
- Je Chante Ung Chant*. "About." <http://jechante.exeter.ac.uk/archive/about.xql?section=introduction>.
- Johns Hopkins University. "Commitment to our Communities" in Johns Hopkins University Progress Report 2015. <http://10x2020progress.jhu.edu/priorities/commitment-to-communities/>
- Johns Hopkins University & Health System. "HopkinsLocal." <http://hopkinslocal.jhu.edu/>
- Leo, Domenic. "Illuminating the *Voir Dit*." In Guillaume de Machaut, *Le Livre dou Voir Dit*. Edited by R. Barton Palmer. Forthcoming.
- Machaut, Guillaume de. *Prologue*. In *The Fountain of Love (La Fontaine Amoureuse) and Two Other Love Vision Poems*. Edited and translated by R. Barton Palmer. New York: Garland, 1993.
- . *Le Livre dou Voir Dit (The Book of the True Poem)*. Edited by Daniel Leech-Wilkinson, translated by R. Barton Palmer. New York: Garland, 1998.
- McGrady, Deborah. *Controlling Readers: Guillaume de Machaut and his Late Medieval Audience*. Toronto: University of Toronto Press, 2006.
- Moi, Toril. *Revolution of the Ordinary: Literary Studies after Wittgenstein, Austin, and Cavell*. Chicago: University of Chicago Press, 2017.
- Mowbray, John, and Edward H. Bouton. *Roland Park Company Records. 1865–1970, Bulk 1891–1960: 1970*. <http://aspace.library.jhu.edu/repositories/3/resources/533>.

- Nichols, Stephen G. "Introduction: Philology in a Manuscript Culture," in "The New Philology." Special issue, *Speculum* 65 (1990): 1–10.
- Palmer, R. Barton. "Guillaume de Machaut and the Classical Tradition: Individual Talent and (Un)communal Tradition." In *A Companion to Guillaume de Machaut*, edited by Deborah McGrady and Jennifer Bain, 241–60. Leiden: Brill, 2012.
- . "The Metafictional Machaut: Self-Reflexivity and Self-Mediation in the Two Judgment Poems." *Studies in the Literary Imagination* 20 (1987): 23–39.
- Plumley, Yolanda. *The Art of Grafted Song: Citation and Allusion in the Age of Machaut*. Oxford: Oxford University Press, 2013.
- Stinson, Timothy L. "Makers of the Mind: Authorial Intention, Editorial Practice, and *The Siege of Jerusalem*." *Yearbook of Langland Studies* 24 (2010): 39–62.
- Text Encoding Initiative. "TEI: Text Encoding Initiative." www.tei-c.org/index.xml
- Timpanaro, Sebastian. *The Genesis of Lachmann's Method*. Edited and translated by Glenn W. Most. Chicago: University of Chicago Press, 2005.
- Underwood, Ted. "Theorizing Research Practices We Forgot to Theorize Twenty Years Ago." *Representations* 127 (2014): 64–72. doi:10.1525/rep.2014.127.1.64.
- Williams, Sarah Jane. "An Author's Role in Fourteenth-Century Book Production." In *Pseudo-Autobiography in the Fourteenth Century: Juan Ruiz, Guillaume de Machaut, Jean Froissart, and Geoffrey Chaucer*, edited by Laurence De Looze. Gainesville: University of Florida Press, 1997.

Chapter 7

OF DINOSAURS AND DWARVES: MOVING ON FROM MOUVANCE IN DIGITAL EDITIONS

TIMOTHY L. STINSON

STUDIES OF *TRANSLATIO* typically focus on transformations of texts that occurred in the context of medieval practices of copying and authorship, including compiling, parsing, and “translating” languages, as well as on authorial efforts to convey *authoritas*. In recent centuries, however, most critical interactions with medieval literary works and their authors have been experienced in and mediated through printed representations of these texts. In creating and using print texts, whether critical editions or mass-market paperbacks, editors and critics have introduced new acts of *translatio* that have remained fairly invisible to the critical lens. Such print-era practices as typography, modernized spelling, footnotes, glossaries, a focus on authorial intention, and a preference for a single, authoritative text rather than the simultaneous presence of competing witnesses, have profoundly shaped conceptions of medieval authorship and textuality and coloured the way we understand, read, and teach medieval literature.

When we discuss medieval *authoritas*, and particularly the authoring of vernacular literature, we cannot do so without recourse to acts of *translatio*. A. J. Minnis has remarked that “[t]he work of an *auctor* was a book worth reading; a book worth reading had to be the work of an *auctor*. No ‘modern’ writer could decently be called an *auctor* in a period in which men saw themselves as dwarfs standing on the shoulders of giants, *i.e.* the ‘ancients.’”¹ The central image Minnis uses may be traced back at least as far as Bernard of Chartres, who reputedly stated, “we see more and farther than our predecessors, not because we have keener vision or greater height, but because we are lifted up and borne aloft on their gigantic stature.”² This metaphor presents a mixed opinion of the dwarves, who simultaneously see more—and farther—but are less capable than their predecessors, a feeling not unfamiliar to many editors today. And what those dwarves were doing when they set quill to parchment constituted acts of *translatio* such as translating, compiling, and synthesizing. At times the goal of these acts was to relay faithfully the *authoritas* of some previous giant. But we also find in these acts of *translatio* medieval processes of authoring literary works, processes that are often far removed from our own notions of what it means to be an author.

It has been noted, reasonably enough, that we are the heirs of Romantic ideals of the author as inspired, solitary genius. And we have also come to view authorship in another, often simultaneous, light as a legal position, *i.e.*, that person who holds the copyright over a creation of the mind, over intellectual property.³ The printed codex,

1 Minnis, *Medieval Theory of Authorship*, 12.

2 Bernard of Chartres in John of Salisbury, *Metalogicon* 3.4, 167.

3 See, for example Wogan-Browne, *The Idea of the Vernacular*, 4.

until at least very recently our primary technological means of containing, shaping, and disseminating authors and their works, played a crucial formative role in the emergence of such ideas; copyright and authorship as we understand them were byproducts of the print era and are not features of manuscript culture. It is no surprise, then, that printed codices have in many ways not been suitable vehicles for accurate reflections of medieval authorial acts and roles, a fact that is in some ways obvious yet one that has attracted insufficient critical reflection. The technology of print joined forces with editorial theory and practice to bring us printed versions of medieval literature whose guiding principle was a quest after a single, “correct” version of a text, a “pure” version and one that conformed most closely to the intentions of an author. As Ralph Hanna notes, “[g]enerally speaking, until the past twenty years, conventional wisdom saw that canonical poets must come pristine: their study can only advance through a fixed text, a verbal icon clearly associable with authorial genius. In such a process, possibly diverse evidence needs to be suppressed and the whole reduced to a single unquestionably authoritative sea of unmarked text.”⁴ Print editions usually either elevate one scribal copy (the rationale and method of best text editions) or present a conflated text created from the best available medieval readings, (i.e., the eclectic edition).

Both approaches of necessity suppress the readings of most manuscript copies of a given work, the best text edition by selecting only one manuscript and the eclectic edition by suppressing those variants not ignored altogether into apparatus and appendices. The more manuscripts there are, the more that are suppressed, and of course this results not in a medieval text but in a modern approximation of one. This is the case, for example, in critical editions of the fourteenth-century alliterative poem *Siege of Jerusalem*, including the Early English Text Society edition that Hanna co-edited with David Lawton, which use Bodleian Library, MS Laud Misc. 656 (L) as a base text. Because L is the sole witness in the *alpha* tradition of the poem and sits alone on one side of a bifurcated stemma, the entire *beta* tradition of the poem, which is that found in eight out of nine surviving copies, is excluded. The goal of recovering the intentions of an author, which is the usual goal of an eclectic edition and also a sentiment that undergirds much of the philosophy behind best text editing, is at odds with notions of authorship and authority found in many medieval manuscript contexts. The “fixed text ... associable with authorial genius” is an expectation of the print era, a notion whose assumptions are clearly etched in everything from our copyright laws to our library cataloguing schemas. And it is a concept that was itself shaped, along with our expectations as readers and consumers of texts, by the print medium, a medium that handles a fixed, “authoritative sea of unmarked text” quite well, but that cannot account for the multiple, overlapping activities of *translatio* and notions of *auctoritas* found in scribal manuscript culture.

Textual Variance and Instability

I have argued elsewhere that textual variance and instability is a hallmark of medieval literary texts, at least as they survive in manuscript copies. My assertion has been (and

⁴ Hanna, *Pursuing History*, 4.

remains) that the *variance*—and variants—produced in this process are vital aspects of medieval literary traditions that can be much better handled by digital media than by printed codices. I have endeavoured to demonstrate this both in argument⁵ and in practice in my work on the *Piers Plowman* Electronic Archive (PPEA) and the *Siege of Jerusalem* Electronic Archive (SJEA). Of course the first part of this argument—that printed books are insufficient for the task—is not new, and I will revisit two well-travelled touchstones in this debate. The first is Bernard Cerquiglini's *Éloge de la variante: Histoire critique de la philologie*, first published in 1989 and republished a decade later in a translation by Betsy Wing as *In Praise of the Variant: A Critical History of Philology*.⁶ The second is the January 1990 issue of *Speculum*, which was edited by Stephen G. Nichols and was devoted to the topic of New Philology. Both of these attracted their fair share of energetic detractors, although the furor has by now died down considerably. Prior to these works was Paul Zumthor's *Essai de poétique médiévale*; Zumthor used the term *mouvance* to refer to the high degree of mobility and change that occurred during the transmission of medieval literary texts. Cerquiglini's work is in part a sequel to Zumthor's, and his conception of *variance* owes much to Zumthor's *mouvance*. But whereas Zumthor was particularly interested in the relationship of orality and authorial anonymity to the mutability of texts, Cerquiglini focused on variants found in written medieval texts, and thus relates more closely to the subject of this essay.

Textual variance is intimately connected to the acts of *translatio* discussed above. As Nichols notes, the mechanical press played a large role in the "movement away from the multiplicity and variance of a manuscript culture, thereby rejecting, at the same time, the representation of the past which went along with medieval manuscript culture: adaptation or *translatio*, the continual rewriting of past works in a variety of versions, a supplementation rather than faithful imitation."⁷ The combination of multiple roles of *auctoritas* and *translatio*, which I have mentioned separately above, in practice naturally overlapped and often occurred in simultaneity (e.g., a scribe's copying of a poet redacting another poet's work), which resulted in a polyvocal text. Any given manuscript is a site of multiple intersections of *auctoritas* manifested through acts of *translatio*. Each manuscript then, is a polyvocal literary performance. At the simplest possible level of medieval transmission of vernacular texts—that of one scribe copying from one exemplar—we have still such intersections, for both the scribe and the exemplar are sites and sources of authority, and most medieval manuscripts quite evidently show such negotiations. But of course things usually were not that simple. Take for example, *Piers Plowman*, the transmission of which is famously complex, with more than fifty manuscripts and three or more textual traditions. And, unlike *Piers*, many vernacular texts of this time relied heavily on source texts as well, so that, for example, in the *Siege of Jerusalem*, the poet simultaneously synthesizes five sources, translates from French and Latin to English, and renders prose sources as long line alliterative verse.

⁵ See, for example, Stinson, "Makeres of the Mind."

⁶ Quotations in this essay are from Wing's translation.

⁷ Nichols, "Introduction," 3.

"Continual rewriting" was the norm during the copying of medieval vernacular texts (although such rewriting was to be avoided as much as possible in certain cases, as when transmitting sacred texts). Poets were often engaged in their own acts of rewriting and redaction. And of course the very business of scribes was "rewriting." Sometimes this was simply the mundane business of copying in a more or less word-for-word fashion, although even this almost inevitably involved accidental reworking of texts through the introduction of mistakes, and frequently involved variant spellings (sometimes due to a difference in dialect) as well as efforts to correct mistakes introduced by previous copyists. But often the scribal copyists themselves performed authorial roles such as adding or deleting passages, editing for clarity or brevity, or introducing commentary and asides. All of this makes for literary texts that were typically both varied and in some way multiple.

From the beginning, print technologies have not been able to capture or retransmit such multiplicity and variance. Or perhaps it would be more accurate to assert that print usually *has not* captured or transmitted such variance. One could, presumably, publish in print format full transcriptions and photo facsimiles of all extant manuscripts of the *Canterbury Tales*. But as welcome as this would be to some scholars, it is, from a business perspective, a recipe for financial ruin and would result in a set of volumes that would be unwieldy to use. But more to the point are the expectations created by this movement away from multiplicity and variance and towards the "authoritative sea of unmarked text" that begins with print. Our very expectations of what a text should be have been shaped by print technologies, and this has caused us to see competing roles of *auctoritas* as well as the acts of *translatio* from which they arise to be problems to sort through and choose among in our efforts to arrive at and print a text. We collapse these roles and activities until they are fairly singular—an authoritative text by *an* author—and until they are suitable for printing. In such versions, multiplicity and textual variance usually take the form of footnotes and appendices that make a good-faith effort to report to informed readers those very places where an editor or critic finds that multiple readings and scribes, translators, or authors push back against such collapse, muddying the waters of singularity.

I am returning to Cerquiglini and New Philology not to glean the fields of those debates, which have been picked over thoroughly by this point, but rather to assess how we might productively move forward in the editing of Middle English literary texts in digital environments. In particular, there are three points that I wish to make:

- 1) Cerquiglini's notion of *variance* has a paradoxical uniformity that does not match the reality of medieval literature broadly, and certainly is an ill fit for Middle English literary traditions.
- 2) There are important differences between Middle English and Old French literary traditions that need to be considered in applying the work of Zumthor, Cerquiglini, Nichols, and others to our field.
- 3) Cerquiglini and New Philologists advocate the use of digital media, but do not explain in any detail what such applications should look like.

Cerquiglini's Invariable variance

In the fourth chapter of *In Praise of the Variant*, "Gaston Paris and the Dinosaurs," Cerquiglini attacks the empirical and positivist approaches of earlier textual critics. He singles out Gaston Paris, who "was suspicious of variants," and Bédier, "who loved them, to the point of editing them exclusively," but who provided "no picture of the intrinsic variance that is medieval writing, only snapshots, which are, of course, preferable to illusory reconstruction."⁸ Throughout the chapter Cerquiglini maintains an extended comparison of philologists to palaeontologists who "took on the task of reconstructing the primordial and perfect *Ursprache*," and adopted "a rational method of classifying manuscripts which promoted the reconstruction of the archetype *Urtext*."⁹ In this same vein, Cerquiglini parallels Bédier with Georges Cuvier, one of the founders of palaeontology who is remembered particularly for helping to found vertebrate palaeontology as a field of study and for being among the first, if not the first, to argue convincingly that many previous species had been subject to mass extinction. As explained by Cerquiglini, Cuvier was a proponent of "the thesis of universal catastrophe that supposedly destroyed all species then alive, so that we no longer have any knowledge of them except through scattered fragments, which, through patient application of comparative anatomy, can be pieced together by the naturalist."¹⁰ Moreover, "Gaston Paris's philology was based on much the same thesis: the copy is a catastrophe that displaces and destroys the original work so that all that remains to us now are (Paris could have repeated Cuvier's words) 'isolated bones thrown helter-skelter, almost always broken and reduced to fragments.'"¹¹ This is an apt and not an unthoughtful comparison. And while it is clearly meant to be a negative one, it is easy to imagine a textual critic engaged in creating a critical edition accepting this as a positive account of her project: from the broken textual remains of numerous individual witnesses scattered here and there, we are able to build an approximation of the lost whole that is fuller and more comprehensible than any single witness (i.e., fossil) by itself. The further implication that Paris himself was an outmoded dinosaur is, of course, less amenable to being taken in a positive light; we might join Evelyn Birge Vitz in wondering "why all this dredging up today of the sins of Philology in the nineteenth century? (Why this obsession with Gaston Paris?)" As Vitz notes, many disciplines at that time had "positivistic and scientific pretensions," but most seem to have been forgiven their awkward adolescences.¹²

In contrast to the fossilized approaches of his nineteenth-century predecessors, Cerquiglini proposes *variance*, in the form of variants found in medieval manuscripts and in his own conception of texts and editing. A central irony is that Cerquiglini's notion

⁸ Cerquiglini, *In Praise of the Variant*, 70–71.

⁹ Cerquiglini, *In Praise of the Variant*, 48.

¹⁰ Cerquiglini, *In Praise of the Variant*, 69.

¹¹ Cerquiglini, *In Praise of the Variant*. Cerquiglini here quotes from Cuvier, *Discours sur les révolutions de la surface du globe*, 53.

¹² Vitz, "A Renewed Philology," 74.

of *variance* is itself fixed and somewhat fossilized. While admitting that variants come in many forms, his work suggests that *variance* is a single unchanging characteristic of medieval manuscripts, and that variants are invariable in the way they are generated and transmitted within vernacular literary manuscripts. This is of particular importance due to the centrality of *variance* to his argument and to his conception of medieval textuality. He argues, for example, that “[i]n the Middle Ages the literary work was a variable,”¹³ that “instability of medieval works in the vernacular is a clear illustration of what is particular to both the written manuscript and, more generally, scribal culture,”¹⁴ and that “[v]ariance is the main characteristic of a work in the medieval vernacular; a concrete difference at the very basis of this object,”¹⁵ *inter alia*. Keith Busby, in a trenchant but reasonably argued response to Cerquiglini, observes that such an approach is hardly possible in one of the most common environments in which we find vernacular literary works, namely as a survivor found in a sole manuscript copy:

A work preserved in a single manuscript can hardly exhibit *variance*, although historically, it might well have done if other copies had existed. The single-manuscript text is, in fact, a useful counter-example to Cerquiglini’s argument in more ways than one, for to us it *is* the text, just as most copies *were* the text for those who read them or had them read in the Middle Ages.... Our ability to compare all extant manuscripts of a work, to perceive in outline and in detail the *mouvance* and *variance* of the medieval text, is therefore as anachronistic as Cerquiglini claims the canonized, printed text to be.”¹⁶

I would add to those objections that *variance* is itself too variable to be discussed as a single feature of all medieval literary texts. Some variants in written texts do indeed constitute the relatively freewheeling type of variety described by Cerquiglini as *variance*. And this is particularly true with literature that, as Zumthor discusses, was a product of oral transmission. But other variants, and *variance*, are demonstrably simply slips of the mind or pen, or are even the products of medieval scribes attempting to correct a text back to the author’s originally intended text, an act that Cerquiglini decries as an anachronistic modern approach.

Moving on to Middle English

Towards the end of his essay, cited above, Busby entertains the possibility that by “responding so vigorously” to Cerquiglini he has “[t]aken him too seriously, just as others have done,” and that the “*Éloge de la variante* is a whoopee-cushion intended for the collective *derrière* of the French philological establishment.”¹⁷ Perhaps Busby

¹³ Cerquiglini, *In Praise of the Variant*, 33.

¹⁴ Cerquiglini, *In Praise of the Variant*, 34.

¹⁵ Cerquiglini, *In Praise of the Variant*, 37–38.

¹⁶ Busby, “The Politics of Textual Criticism,” 32–33.

¹⁷ Busby, “The Politics of Textual Criticism,” 44.

takes things too far here, but this characterization points to two important aspects of Cerquiglini's work. First, there is a great deal of wit and a clear sense of play in the book, and we would be remiss to read or respond to it without acknowledging this. Neither should we discount the fact that there is a serious project in this slim volume, however, as it identifies real and substantial problems with the ways in which medieval literary texts have been transmitted through printed editions. Second, Cerquiglini concerns himself primarily with French literature and addresses scholars and editors in that field. The same was true of Zumthor, of Nichols, and of most of the contributors to the "New Philology" issue of *Speculum*. This is of course not a limitation inherent in any of these scholars or their work. Quite the contrary; Middle English literary studies and editorial theory have responded slowly or not at all to the issues of *mouvance* and *variance* and how these might best be handled in critical editions, which is why we need to borrow from the conversations of those working on Old French texts to move forward in our own field. This has begun to change with the publication of volumes such as Vincent Gillespie and Anne Hudson's *Probable Truth: Editing Medieval Texts From Britain in the Twenty-First Century* and, well before that, Tim William Machan's *Textual Criticism and Middle English Texts*. But this work has only begun. The need to translate these ideas into theory and practice within the field of Middle English literary studies and editorial practice remains.

I argue above that Cerquiglini's conception of *variance* is, ironically, too fixed. He seems to imagine the creation of variants as a vital tradition of recomposition by scribes that places them in a more or less authorial role, and he denigrates the notion that *variance* is in fact sometimes nothing more than careless corruption of an authorial original. Cerquiglini engages with Middle English texts and editorial approaches only briefly and obliquely:

The theory that the copy represents degeneration presupposes a flawless original; the author has no right to any lapsus. Similarly, the idea that language becomes degraded implies an impeccable origin: the author has no right to bad language either, or to dreadful puns, or, indeed, to the diversity of his way of speaking. By subscribing to this, philology surreptitiously annexed a literary theory—the theory of the genius. By magnifying a transcendent author it tied an authoritarian theory of the subject (the master of the sense as well as of the signifier expressing it) to the notions of origin and textual stability. The author, great by definition, and unique, the most pre- of pre-production by the unity of his conception, the opacity of his work (the argument of the *lectio difficilior*), and the quality of his language, stood in sharp contrast to scribal diversity, ignorant and purposeless, which pluralized the work, trivialized its expression, and impoverished its language.¹⁸

A footnote attached to the second sentence of this passage points to an essay by Lee Patterson, "The Logic of Textual Criticism and the Way of the Genius: The Kane-Donaldson

18 Cerquiglini, *In Praise of the Variant*, 61.

Piers Plowman in Historical Perspective.”¹⁹ This, combined with the mention of “the argument of the *lectio difficilior*,” signals that Kane-Donaldson and their work on *Piers Plowman* are the targets here. Some of the problems with this paragraph are self-evident. Clearly an original text that is not impeccable, that is chock full of bad language and dreadful puns, is subject to the same vicissitudes of time and scribal transmission as a putative impeccable text by an impeccable author. Nor can modern editorial decisions anachronistically alter the rights of dead authors, although his point is taken. And only a very careless reader could believe that Kane or Donaldson understood *Piers Plowman*, of all texts, to be a model of textual stability or unity of conception. They were struggling against quite opposite realities. Finally, we see here once again the tendency to view *variance* as invariable. The truth is that a considerable amount of the scribal activity—the *variance*—found in in *Piers Plowman* manuscripts did indeed pluralize and impoverish its language.

The discussions of *mouvance* or *variance* in medieval French texts raise issues of importance to Middle English texts as well, yet they are not easily applicable to many medieval English literary texts. The best fit, not surprisingly, would be for English romances that, like *Sir Orfeo*, to take but one example, survive in multiple substantially different manuscript versions, and have roots in French romance tradition. Here we do indeed find both *mouvance* and *variance* as Zumthor and Cerquiglini understand them. Of course many Middle English romances survive in unique manuscript copies, which returns us to Busby’s point that “a work preserved in a single manuscript can hardly exhibit *variance*.” And for many other Middle English texts, the concept hardly applies, a point that has been made forcefully by Derek Pearsall:

But it is clear that the idea of *mouvance* is irrelevant to written texts that had no previous oral existence, where there is a single author, and where the manuscript variations can be analysed according to the usual strategies for determining authenticity. Nevertheless, we are often told that all texts in the manuscript era, since every manuscript is unique, participate in a state of continuous fluidity and chaotic particularity, so that editing in the usual sense of the word, that is the determination of the relative authenticity of readings, is impossible.²⁰

Finally, it should be added that the variants produced by many Middle English scribes seem clearly to be the products of efforts on the parts of those scribes, not entirely dissimilar from those of modern editors, to transmit a corrected text that is as accurate as possible. There is ample evidence of this within the corpus of surviving *Piers* manuscripts. Following extensive work with a variety of copies of the poem in the course of working on his critical text of Bx,²¹ Thorlac Turville-Petre notes that while we “have grown so used to the notion of textual variability and error in *Piers Plowman* that it is

19 Patterson, “The Logic of Textual Criticism and the Way of the Genius.”

20 Pearsall, “Variants vs Variance,” 201–202.

21 *Piers Plowman* Electronic Archive, vol. 9: *The B-Text Archetype*.

easy to come away with the impression that Langland's scribes were incompetent and careless," he has conversely been impressed by "the very considerable trouble that the scribes took to get their text right, if possible even righter than their exemplar."²² We see, then, that within Middle English manuscripts *variance* is, well, highly variable. It is not possible to discuss the concept as a single type of scribal behaviour that can be in turn dealt with—or mishandled—in a uniform manner. And the same thing is surely true of Old French and other vernacular literatures as well, although I will leave those arguments to specialists in the respective fields.

Moving Online

Cerquiglini concludes his work by stating that the computer is the "obvious solution" to the limitations and distortions of "the two-dimensional space of the printed page" that are results of presenting medieval literary works in printed editions. He advocates a new way of presenting written texts that he terms *écranique*, translated here as "screenic":

Always in the process of further development and refinement by technicians, the screen is simultaneously dialogic (it offers a constant interaction between the user and the screen) and multidimensional (through the use of "windows," it allows one to bring together and consult information belonging to separate entities). Utilizing these two qualities, one can conceive of the type of edition that would result from this assembling of separate entities represented by codices. In such an edition medieval works would no longer be subjected to the two-dimensional and closed structure of the printed page because a diskette accommodates varied textual masses, which the reader consults by making them appear in different ways on a computer screen.²³

While some of the details have of course changed since the 1980s—we now tend to use web-based editions rather than diskettes—Cerquiglini describes here just the sorts of electronic editions of medieval texts that would appear over the next couple of decades. Both the PPEA and the SJEa feature drop menus that allow users of the editions to move between four views of each text: Diplomatic, Critical, Scribal, and All Tags (Figure 7.1).²⁴ Texts may also be searched, compared, and displayed in a variety of other ways. The critical edition of Bx published by the PPEA, for example, allows users to move between documentary texts and editions and the critical text by providing an interactive apparatus. Clicking on a line number in the left margin of the critical edition opens the apparatus in a window that contains full lines from ten of the most important B manuscripts (Figure 7.2).

²² Turville-Petre, "Putting it Right," 41.

²³ Cerquiglini, *In Praise of the Variant*, 79.

²⁴ For a description of the purpose of each style sheet and the differences between them, see <http://siegeofjerusalem.org/index.html?page=use>



Figure 7.1. Four views of the text available on the *Siege of Jerusalem* Electronic Archive.

Bx.1.36 [KD.1.34]	For-þi drede delitable drynke · and þow shalt do þe bettere
XL.1.36 [KD.1.34]	For-þi drede delitable drynke · and þow shalt do þe bettere
M.1.36 [KD.1.34]	For-thi drede delitable drinke : and þou shalt do þe bett(er)re
Cr.1.1.35 [KD.1.34]	Forthy drede delectable drinke , & thou shalt do þe better
W.1.36 [KD.1.34]	¶ For-þi dred delitable drynke . and þow shalt do þe bettre
Hm.1.35 [KD.1.34]	forþi drede delytable drynke : and thu schalt do the bettre
C.1.34 [KD.1.34]	For-thy drede delitable drynke · and thou schalt do þe better
G.2.36 [KD.1.34]	forthy drede deylf(e)ctable LG dryn(ke) & þou shalt do þe better
O.1.34 [KD.1.34]	For-þi drede delitable drynk : & þou schalt do þe better
R.1.36 [KD.1.34]	¶ Forthi drede delytable drynke : and þow schalt do þe bettre .
F.2.36 [KD.1.34]	For-þy : drede delytable drynk / & þou shalt do þe bettre.

Figure 7.2. Apparatus of Bx, *Piers Plowman* Electronic Archive.

I agree with Cerquiglini regarding the capacity of electronic editions to be flexible and multi-dimensional. But another fundamental affordance of such editions is their ability to escape many of the dichotomous choices mandated by printed editions. As a technology, print does not lend itself easily to multiple views of the same text or information, at least not on the same plane. There are limited ways in which one might achieve this, such as parallel text editions, but for the most part I am simply making a commonsense claim; one need only to consider the amount of space it would take to

provide the ten-line apparatus for the PPEA edition of Bx in print, and how unwieldy such a thing would be to use, to see this. Printing something one way, in short, has the practical consequence of ruling out alternatives in most instances. While digital editions have moved us beyond the technological constraints that mandated either/or decisions, they have not yet prompted us beyond either/or thinking about texts and how we should edit them. And this is as true of Cerquiglini's work as it is of his imagined character "Mr. Procrustes, Philologist" (the title of the second chapter of his book).

One such dichotomy is the debate surrounding best text versus eclectic editing. In the context of editing Old French texts, this debate might be summed up as Bédierist best text editing versus Lachmannian genetic editing, which was "introduced with great fanfare by Gaston Paris at the end of the 1860s into the bosom of French medievalism."²⁵ This dichotomy is of course not unfamiliar to textual critics working on Middle English. As Machan notes, for many "the advocated method has increasingly become essentially diplomatic transcription," while others "have fully embraced humanist textual criticism and therefore humanist assumptions."²⁶ (So as not to misrepresent the nuances of Machan's argument, I should acknowledge here that he thoughtfully argues that even the best-text approaches are, like Bédier, "nonetheless informed by unassessed humanist notions of the authoritative, lexical work"). Machan profitably explores the ways in which approaches to medieval texts are undergirded—and distorted—by anachronistic humanist conceptions. We should add to this the recognition that technology has also been a significant driving force in how we conceive editing, and in particular that print both enables and constrains us in terms of how we represent and transmit medieval texts. Usually presented as antithetical theoretical and practical approaches to texts, best text versus eclectic editing may also be understood as a dichotomous practical choice that someone preparing a printed edition of a medieval literary text would need to make. A variety of theoretical factors might play into such a decision, but in the end it is difficult for practical reasons, including space, funding, and usability, to create a printed book that simultaneously embraces the benefits of both approaches. But of course this is entirely possible with digital media, as the PPEA makes clear. Burrow and Turville-Petre's critical edition of Bx dynamically incorporates documentary editions of manuscript witnesses in the apparatus, as discussed above. And the documentary editions themselves handle the texts in flexible ways, as the menu of views allows the user to privilege the document itself (diplomatic view), the intentions of the scribe (scribal view), or those of the editors (critical view), and more (the user may also view all tags at once, work directly with the xml-encoded transcriptions, or bypass editorial intervention by working directly from digital images of the respective folios). In calling for digital editions, then, there is no need to denigrate either of these supposedly antithetical approaches, or even to choose between them, as they can in fact be harmonious parts of a larger edition or digital editorial project such as the PPEA.

²⁵ Cerquiglini, *In Praise of the Variant*, 51.

²⁶ Machan, *Textual Criticism and Middle English Texts*, 61–62.

A second example of either/or thinking that we might eclipse with digital editions is the paradoxically inflexible approach that Cerquiglini takes toward *variance* and variants. As discussed above, variants occur in many forms. The editor of *Sir Orfeo*, with the divergent readings typical of romances surviving in multiple manuscripts, faces one set of problems, while the editor of a romance surviving in a single, evidently corrupt copy faces another. And these are a far cry from texts like Chaucer's major texts, where "the archetype is so reliable, whether because of the excellence of the copyists or because of authorial supervision, that it can be largely trusted."²⁷ The alliterative *Siege of Jerusalem* is also an example of a work where establishing a critical text is an entirely reasonable goal, as Hanna and Lawton have demonstrated in their EETS edition. Because the poet closely translated from French and Latin sources and the editors have other criteria such as metricality, sense, and completeness to guide them, it is reasonable in most cases to propose authorial readings with confidence.

Piers Plowman is a poem that always manages to be an outlier, and so it is here. It is in some ways, like *Siege of Jerusalem* or the major Chaucer poems, an attractive candidate for critical editions. But while those poems have their own sets of complexities to occupy editors, they are in no way as complex as the surviving corpus of *Piers* texts and the editorial cruces that they famously pose. This is due in large part, of course, to the fact that there are three versions of the poem and a large number of surviving manuscripts (although fewer than Chaucer manuscripts) that contain among them significant differences in readings, shifts in exemplar, splices of two or more versions of the poem, and other challenges and eccentricities. On the one hand we have scribes like those of Huntington Library MS Hm 128 and British Library Additional MS 35287, the two B texts that are the subject of Turville-Petre's essay, cited earlier, that demonstrate the pains that some copyists took in transmitting, or even improving upon, their exemplars. On the other hand we have manuscripts like Huntington Library MS Hm 143, a *Piers* C text that, as Michael Calabrese has demonstrated, features corrections and erasures that "provoke important questions about our understanding of textuality, authority, and correction, as well as about the charged topic of the relationship between Langland and Wyclifism."²⁸ Here variants and *variance* do not mean the same thing within the textual tradition of a single poem. The B text copyists that Turville-Petre discusses produced variants that are useful evidence for, among other things, establishing critical and archetypal texts. The C text copyist that Calabrese discusses produces variants, of course, but also a unique, engaged reading of the poem that might more reasonably be called *variance* as Cerquiglini means it. But in either case we can build a digital edition and archive that employs this evidence in a variety of ways. The variants and *variance* are preserved in documentary editions of the individual manuscripts in question.²⁹ These

²⁷ Pearsall, "Variants vs Variance," 202.

²⁸ Calabrese, "[Piers] the [Plowman]," 171. Calabrese and Turville-Petre's essays were discussed together in a similar context in "The *Piers Plowman* Electronic Archive on the Web: An Introduction," an essay that I co-authored with Jim Knowles.

²⁹ The editions of the B texts have been published as part of the *Piers Plowman* Electronic Archive at <http://piers.chass.ncsu.edu/texts/Hm> and <http://piers.chass.ncsu.edu/texts/M>. The C text, Hm 143, is forthcoming from PPEA.

documentary editions might include in them the option of increasing or decreasing the level of authority granted to the scribe, the editor, or the document itself. And as Burrow and Turville-Petre have demonstrated, a wide range of variants from documentary texts may serve as the basis of a critical text without that evidence needing to be merely ignored or suppressed into apparatus.

While critical editions reflecting authorial intention might on the surface of things seem to conflict with Cerquiglini's vision, there can be no doubt that in fact the digital editions discussed in this essay answer his call for editions in which "medieval works would no longer be subjected to the two-dimensional and closed structure of the printed page." Yet they do so in a way that avoids overly restrictive, dichotomous ways of viewing editorial approaches and the treatment of variants, or *variance*. Editors today stand in a vexed relationship with print. There is on the one hand a debt of gratitude to the medium that for centuries has allowed editorial theory and practice to exist and to develop. It has been our primary means both of reading medieval texts and of reading about them. Increasingly, however, it has begun to appear both a fortuitous and problematic intermediary between manuscripts and digital editions. In order to move beyond the limitations of printed editions and to realize the potential (and, yes, limitations) inherent in digital technologies, we must first have a clear vision of the habits and mindsets inculcated by producing and consuming printed versions of medieval texts that were themselves products of manuscript culture. A good first step towards this is to acknowledge and move beyond dichotomies that are themselves merely byproducts of print-era practices and texts and not inherent features of medieval textuality.

Bibliography

- Busby, Keith. "Variance and the Politics of Textual Criticism." In *Towards a Synthesis? Essays on the New Philology*, edited by Keith Busby, 29–45. Amsterdam: Rodopi, 1993.
- Calabrese, Michael. "[Piers] the [Plowman]: The Corrections, Interventions, and Erasures in Huntington MS Hm 143 (X)." *Yearbook of Langland Studies* 19 (2005): 169–99.
- Cerquiglini, Bernard. *Éloge de la variante: Histoire critique de la philologie*. Paris: Éditions du Seuil, 1989.
- . *In Praise of the Variant: A Critical History of Philology*. Translated by Betsy Wing. Baltimore: Johns Hopkins University Press, 1999.
- Cuvier, Georges. *Discours sur les révolutions de la surface du globe, et sur les changements qu'elles ont produits dans le règne animal*. Paris: Schleicher, 1809.
- Gillespie, Vincent, and Anne Hudson, eds. *Probable Truth: Editing Medieval Texts from Britain in the Twenty-First Century*. Turnhout: Brepols, 2013.
- Hanna, Ralph. *Pursuing History: Middle English Manuscripts and Their Texts*. Stanford: Stanford University Press, 1996.
- Hanna, Ralph, and David Lawton, eds. *The Siege of Jerusalem*. Early English Text Society 320. Oxford: Oxford University Press, 2003.
- John of Salisbury, *The Metalogicon: A Twelfth-Century Defense of the Verbal and Logical Arts of the Trivium*. Translated by Daniel McGarry. Berkeley: University of California Press, 1955.

- Knowles, Jim, and Timothy L. Stinson. "The *Piers Plowman* Electronic Archive on the Web: An Introduction." *Yearbook of Langland Studies* 28 (2014): 225–38.
- Machan, Tim William. *Textual Criticism and Middle English Texts*. Charlottesville: University of Virginia Press, 1994.
- Minnis, A. J. *Medieval Theory of Authorship*. 2nd ed. Philadelphia: University of Pennsylvania Press, 1988.
- Nichols, Stephen G. "Introduction: Philology in a Manuscript Culture," in "The New Philology" Special issue, *Speculum* 65 (1990): 1–10.
- Patterson, Lee. "The Logic of Textual Criticism and the Way of the Genius: The Kane-Donaldson *Piers Plowman* in Historical Perspective." In *Textual Criticism and Literary Interpretation*, edited by Jerome J. McGann, 55–91. Chicago: University of Chicago Press, 1985.
- Pearsall, Derek. "Variants vs Variance." In *Probable Truth: Editing Medieval Texts from Britain in the Twenty-First Century*, edited by Vincent Gillespie and Anne Hudson, 197–205. Turnhout: Brepols, 2013.
- Piers Plowman* Electronic Archive. Society for Early English & Norse Electronic Texts. <http://piers.chass.ncsu.edu>.
- . Vol. 5: *London, British Library, MS Additional 35287 (M)*. Edited by Eric Eliason, Hoyt N. Duggan, and Thorlac Turville-Petre. SEENET Series A.7. The Society for Early English and Norse Electronic Texts, 2014. <http://piers.chass.ncsu.edu/texts/M>.
- . Vol. 6: *San Marino, Huntington Library, MS Hm 128 (Hm and Hm2)*. Edited by Michael Calabrese, Hoyt N. Duggan, and Thorlac Turville-Petre. SEENET Series A.9. The Society for Early English and Norse Electronic Texts, 2014. <http://piers.chass.ncsu.edu/texts/Hm>.
- . Vol. 9: *The B-Text Archetype*. Edited by John Burrow and Thorlac Turville-Petre. SEENET Series A.12. The Society for Early English and Norse Electronic Texts, 2014. <http://piers.chass.ncsu.edu/texts/Bx>.
- Siege of Jerusalem* Electronic Archive. Edited by Timothy L. Stinson. Society for Early English and Norse Electronic Texts. <http://siegeofjerusalem.org>.
- Stinson, Timothy L. "Makeres of the Mind: Authorial Intention, Editorial Practice, and *The Siege of Jerusalem*." *Yearbook of Langland Studies* 24 (2010): 39–62.
- Turville-Petre, Thorlac. "Putting it Right: The Corrections of Huntington Library MS Hm 128 and BL Additional MS 35287." *Yearbook of Langland Studies* 16 (2002): 41–65.
- Vitz, Evelyn Birge. "On the Role of a Renewed Philology in the Study of a Manuscript- and Oral-culture." In *Towards a Synthesis? Essays on the New Philology*, edited by Keith Busby, 71–78. Amsterdam: Rodopi, 1993.
- Wogan-Browne, Jocelyn. *The Idea of the Vernacular: An Anthology of Middle English Literary Theory, 1280–1520*. University Park: Pennsylvania State University Press, 1999.
- Zumthor, Paul. *Essai de poétique médiévale*. Paris: Ed. Du Seuil, 1972.

Chapter 8

ADAM SCRIVEYN IN CYBERSPACE: LOSS, LABOUR, IDEOLOGY, AND INFRASTRUCTURE IN INTEROPERABLE REUSE OF DIGITAL MANUSCRIPT METADATA¹

BRIDGET WHEARTY

And for ther is so gret diversite
In Englissh and in writyng of oure tonge,
So prey I God that non myswrite the,
Ne the mysmetre for defaute of tonge;
And red wherso thow be, or elles songe,
That thow be understonde, God I biseche!
But yet to purpos of my rather speche

—Geoffrey Chaucer, *Troilus and Criseyde*²

Once you start to aggregate these resources and combine them in a new context and for a new purpose, you find out, in practical terms, what it means to say that their creators really only envisioned them being processed in their original context.... It's as though the data has suddenly found itself in Union Station in its pajamas: it is not properly dressed for its new environment.

—John Unsworth, "Digital Humanities Centers as Cyberinfrastructure"³

In recent years, the digitization of medieval manuscripts has grown from a few intrepid trickles to a global flood. Driven by twin commitments to preservation and access, major

¹ This chapter is indebted to the Council on Libraries and Information Resources (CLIR) and the many members of Stanford University Libraries' Department of Digital Library Systems and Services who welcomed and mentored me throughout my postdoc there. I am grateful, especially, to Tom Cramer, Tony Navarrete, Bess Sadler, and Laney McGlohon; and to Greta de Groat, Lynn McRae, and Laura Wilsey who generously shared expertise in metadata and crosswalking. Throughout my postdoc, Benjamin Albritton was an invaluable sounding board, patient teacher, and enthusiastic collaborator. While I am deeply grateful to all at CLIR, SUL, and DLSS, my gratitude to him knows no bounds.

² V.1793–1799. All quotations from Chaucer, unless otherwise noted, come from Benson, ed., *The Riverside Chaucer*.

³ Unsworth, "Digital Humanities Centers as Cyberinfrastructure," quoted in Unsworth "Computational Work with Very Large Collections," 4.

digitization efforts dedicated to medieval books have arisen across the planet.⁴ This tide has also given rise to important warnings that digitization is fundamentally transforming scholarly work upon medieval books.⁵ Although “the physical” and “the digital” are frequently presented in opposition, beneath significant disagreements all sides share core convictions. Namely, that medieval books matter, that they must be cared for, preserved, read, and maintained, that these objects’ physical forms—their words, miniatures, margins, fore-edges and bindings—are vitally important to uncovering complex textual meanings, and to recovering the identities, concerns, and desires of the people who made and read these books centuries before us.

Atop these core convictions, there are significant differences between the work that can be done on physical, *analogue* codices and that which can be done using digital medieval manuscripts. In photographs, much of the sensual experience of the codex goes missing: one cannot touch it, hear it, smell the book; one cannot see the particular movement of these specific leaves as they respond to the exact humidity of the weather on the day of the reader’s visit. Although the language of these objections foregrounds an undeniable pleasure in touching the physical object, lack of physical access raises serious scholarly concerns that cannot be dismissed as the “simple” thrill of touching a medieval book.⁶ Collation formulae, for example, are difficult to determine, or verify, when photographs do not fully show an opening’s gutter, and when the only images of the head and tail show the covers pressed firmly closed. (Moreover, there are valid and real concerns that digital manuscripts’ existence might be used to bar desirous researchers from laying hands on the beloved object.⁷) And yet, despite the undeniable benefits of working directly with the physical book, there is also something equally powerful

4 Among many remarkable projects, the work by the Hill Museum and Manuscript Library stands out: both in creating digital preservation copies of the medieval manuscripts of Timbuktu and in ongoing efforts in Iraq and Syria to create access copies of priceless cultural heritage items, many of which are, at the time of writing, unreachable (and increasingly endangered) due to ongoing political campaigns carried out by the Islamic State. It is also worth noting that while medieval manuscripts are not generally physically destroyed in the process of digitizing, more modern texts have been—and in some cases, continue to be—cut apart in order to aid the speed at which they can be digitized. See, for example, Harvard’s “Free the Law” Project, which—treating books of case law as content vessels rather than objects with their own complex histories—slices books’s pages from their spines and covers in order to achieve the admirable goal of a universal database of American law. In these more modern examples, unlike the medieval ones this chapter is dedicated to discussing, the commitments of preservation and access may sometimes be less “twin goals” and more competing teams.

5 Edwards, “Back to the Real?” remains a key thought-piece in these debates. Similar concerns about the ways that digitized objects fail to convey the full materiality of the physical object are raised in Treharne, “Fleshing Out the Text”; Rudy, “Dirty Books”; Echard, “Containing the Book: The Institutional Afterlives of Medieval Manuscripts,” and her coda to *Printing the Middle Ages*.

6 Not that these thrills are ever simple. Nor should that very real visceral pleasure be dismissed.

7 On the practice of providing access to digital images of manuscripts rather than to manuscripts themselves, see Nikolova-Houston and Houston, “Building the Virtual Scriptorium,” 232. On the impact of digitization on reading room use, see Rudy, “Dirty Books,” 30.

about possessing the ability to call up images of the texts we study whenever, and wherever, we wish. Sitting at home in upstate New York, I can pour a glass of wine, or nibble cheese, as I read Huntington Library, HM 286, a deluxe copy of John Lydgate's *Fall of Princes* (ca. 1440–1460) held in southern California—approximately 2700 miles/2445 km from my home. Although I cannot touch the flesh or smell the leaves of the book, the experience of this kind of slow reading is still strange and lovely: the manuscript and I commune together, through a kind of modern image-magic of the glowing screen. And, significantly, our connection cannot be severed by limited research funding or special collections' reading room hours. I stop reading when I want to, and not before.⁸

This sense of digital access as an intimate, one-to-one communion of lone researcher with distant book is founded on a great deal of invisible and undervalued labour. A vast community of curators, photographers, and metadata and remediation specialists make possible my communion with HM 286, even if—as part of their labour—they erase all trace of their interventions. And this community grows infinitely larger, and even more invisible, when it comes to the labour of making digital manuscripts *interoperable*, by which I mean discoverable, shareable, and useable between and across institutions. The longed-for quest object of digital manuscripts may be a single hub where a researcher can search and view the manuscripts of the world, but this marvellous hub cannot exist without serious work on digital manuscripts' descriptive metadata—the machine-readable descriptions that make it possible for researchers to search by writing support, language, contents, subject, author, scribe, artist, location of origin, current repository, etc. Perhaps because we are significantly less interested in the living than the dead, medievalists rarely inquire into the labour of making and maintaining digital manuscripts. But that is our error. For digital books are, undeniably, still objects in their own rights, albeit objects of a very different sort than the “analogue” book.⁹ Furthermore, the frequently invisible work that goes into their creation and maintenance has a profound impact on the images we see on our screens, as well as those that exist but we cannot find.

This chapter explores how invisible decisions governing the creation and sharing of digitized medieval manuscripts exert a profound, albeit unacknowledged, impact on the kind of research manuscript scholars are able to do. In particular, I focus on the use and reuse of manuscripts' descriptions that I participated in as a CLIR Postdoctoral Fellow in Data Curation for Medieval Studies at Stanford University, working as data curator on the Digital Manuscript Index (hereafter DMS-Index), a multi-institution

8 I acknowledge also that this digital access is a privileged and precarious state, dependent on access to electricity, an up-to-date computer, and a stable and powerful internet connection on my end, as well as everything working smoothly on the home institution's end—and this does not always happen.

9 I am indebted to my colleague Kristen Gallant for this reminder, which has long been accepted common sense among librarians and has shaped library-based practices in producing objects' records. This careful, longstanding differentiation between “analogue” books and digital manuscripts in library communities can provide an important model for manuscript scholars seeking language to differentiate between the two, extremely different objects.

experiment working toward that dream of a single search portal through which the digitized manuscripts of the world may be seen.¹⁰ Because this chapter exists at the intersection of library, medievalist, digitization, and digital humanities (DH) cultures, I begin by defining a broad set of key tools and concepts. Next, using the sharing of manuscripts from the Walters Art Museum with DMS-Index as my case study, I explore how the processes of copying and transformation, inevitably, create loss and change. Then, I move on to the Parker on the Web app and the challenges of reusing older digital projects' data made according to earlier best practices, which might no longer rule. Finally, drawing on my experiences with metadata from e-codices, the digital manuscript library of Switzerland, I consider how cataloguing languages and decisions feed into debates over outreach, access, and the problematic monolingualism of digital humanities.

Throughout, this chapter will argue that manuscript scholars and medievalists must take a more book historical approach to digitized texts. In other words, I seek to promote a codicology of the *digital* medieval book. Rather than treating the digital book as a not-quite-good-enough window to the distant physical book upon which one might like to do codicological study, I argue that manuscript scholars and medievalists must engage in rigorous analysis of the digital book as a unique object. We need to uncover the historical, political, social, and economic pressures that shape digitization and the building of digital repositories, because these forces exert as much influence on the shape of digital medieval texts as they did, centuries ago, in the shaping of physical texts. Framed via a rather different terminology, this chapter seeks to promote a deeper sort of information literacy for medievalists and manuscript scholars, by revealing how the information we see when we work with digital manuscripts is structured by all kinds of outside social forces. I do so because, whether approached via codicology or information literacy, it is imperative for medievalists and manuscript scholars to not lose sight of the fact that all of this information and these manuscript descriptions are always based on human decision-making—and are therefore never neutral.

In particular, this chapter seeks to make visible some of digital projects' invisible labour by sharing my narrative of growth as a programmer and digital humanist in the making of the second iteration of DMS-Index. This chapter is in many ways my conversion narrative, or *bildungsroman*, as a digital humanist. When my work on this project began, I was a freshly minted PhD in English, specializing in fifteenth-century English poetry. I had no digital humanities experience, and so—as is often the case for both library work and digital humanities—much of my learning occurred on the job. Thus, part of the argument of this chapter will be to trace a narrative of skills growth. Similar narratives of growth, I contend, are part of the underlying, hidden story of DH projects and this unseen growth can exert a profound, often unacknowledged impact on the shape these final projects can take. Because this is a *bildungsroman*, there will be readers of this chapter who came of age earlier and who therefore are more advanced

10 On DMS-Index, see Foy, "Media Archaeology and the Digital Incunable," 134.

coders than I was at the moment I began my postdoc. I expect these readers will come up with coding solutions that I did not in my first years of DH practice. But that is perhaps the least interesting kind of response I seek to foster.¹¹ A much larger part of this article's argument is that there is value in revealing this inner working—how what was built looks the way it does *because* I was a digital humanist in formation. A medievalist perspective on this digital work reveals that these processes of self-formation are part of a much longer genealogy of text-technologies and transformation than is often acknowledged.¹² Rather than obscure the learning that took place as I copied, or the ways that my improved coding changed what I was able to do with different metadata sets as the project progressed, I seek to reclaim how copyists, in the constant state of forming their own best practices, have always and will always be the story of medieval texts—in digital archives and aggregators, in earlier projects that those aggregators reuse and curate, all the way back to medieval manuscript makers, who were also involved in the work of copying under social, technological, educational, and economic pressures.

Laying Foundations: Definitions for “Metadata” and “Interoperability”

Metadata

A widely used (if maddeningly circular) definition of *metadata* is “data about data,” or “data that describes or gives information about other data,” definitions that frankly do little to clarify what metadata actually *is* and *does*.¹³ More focused communities offer more specific and useful definitions. In academic libraries, the term is “commonly used for any formal scheme of resource description” for all library holdings that must be catalogued and described—both physical books and digital objects.¹⁴ As the National Information Standards Organization (NISO) puts it, metadata is “structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource.”¹⁵ There are three distinct categories of metadata: administrative

11 Adeline Koh's treatment of how certain class, gender, geographical, racial, and ethnic backgrounds support or inhibit early technological uptake in those who go on to become digital humanists is informative for my understanding here, as is work by Eszter Hargittai, Moya Z. Bailey, Chris Bourg, and Bess Sadler. See Koh, “Niceness, Building, and Opening the Genealogy of the Digital Humanities”; Hargittai, “Digital Natives or Digital Naïves?”; Bailey, “All the Digital Humanists Are White, All the Nerds Are Men, but Some of Us Are Brave”; and Sadler and Bourg, “Feminism and the Future of Library Discovery.”

12 Bredehøft's *The Visible Text: Textual Production and Reproduction from Beowulf to Maus* is exemplary in its working putting medieval texts into a much longer history of text and textual transformation, and has strongly influenced my thinking.

13 OED “metadata” (n.); Schmitt, “Towards an Interoperable Scholarly Edition,” 9.

14 NISO, “Understanding Metadata,” 1.

15 NISO, “Understanding Metadata,” 1. See also Hodge, “Metadata Made Simpler.”

metadata,¹⁶ structural metadata, and *descriptive metadata*.¹⁷ This last category is, arguably, most important for researchers of medieval manuscripts and is the focus of this chapter.

A manuscript's bibliographic record and/or catalogue description—which outlines its contents, physical make-up, provenance, etc.—is its descriptive metadata. It is this descriptive metadata that computers query when researchers seek out manuscripts containing works by John Lydgate, or manuscripts in original fifteenth-century limp bindings, or containing miniatures of St. Margaret of Antioch. For a computer to successfully find manuscripts that fit these categories, this information about the book must be not just represented in the digital images of the book but also encoded in the digital manuscript's descriptive metadata.¹⁸

In addition to essential search-and-discovery functionality, descriptive metadata plays an increasingly important role filling the experiential gaps between what can be learned studying the physical book and what can be perceived working with a digital manuscript via a modern screen. As Mohammed Ourabah Souala and Mohamed Hassoun put it:

simply providing images of the manuscript [via digitization] is not sufficient as the images do not convey many important aspects of the manuscript including: codicological description (codex, binding, etc.); paleographic description (handwriting, etc.); manuscript transmission history.¹⁹

Until the time when providing full text versions of the manuscripts is feasible, manuscript cataloguing²⁰ proves to be a realistic solution for providing access.²¹

It is widely recognized that digitized manuscripts are fundamentally lacking when it comes to fully representing the rich complexity of a physical codex. A. S. G. Edwards highlights several particularly important aspects that are difficult to discern via digital images:

it is often difficult for viewers to take in the scale of the object being presented. It is also difficult to discern distinctions between materials such as parchment

16 This in turn can be separated further into rights management metadata and preservation metadata.

17 NISO, "Understanding Metadata," 1.

18 For more on metadata, and how the term's meaning shifts, Proteus-like, from context to context, see Gilliland, *Introduction to Metadata*, and Schmitt, 9.

19 Soualah and Hassoun, "A TEI P5 Manuscript Description Adaptation for Cataloguing Digitized Arabic Manuscripts," 1.

20 The use of "cataloguing" here as a synonym for "metadata" highlights how much descriptive metadata is intimately connected to earlier physical records. For an exploration of how older data continues to shape readers' understanding of the books they see on their screens, see Echard, "Containing the Book," 104–105.

21 Soualah and Hassoun, "A TEI P5 Manuscript Description Adaptation," 2.

and paper, and between different textures of ink. Often we can't tell what the overall structure of the work is like, how many leaves it has and whether it contains any cancel leaves.²²

Elaine Treharne describes challenging her graduate students to determine the size of Cambridge, Corpus Christi College, Matthew Parker Library, MS 391 based solely on digital images from the Parker on the Web.²³ Her students' widely different suggestions (ranging from 15 cm x 7 cm all the way up to 76 cm x 45.7 cm) demonstrate how images alone are insufficient for assessing a book's size and function.²⁴ This is shared as evidence of the limitations of the digital image, all that it cannot successfully convey.

Yet this is also what descriptive metadata is for. Students in my course "Medieval Books in the Digital Age" (Binghamton University, Fall 2015) repeatedly emphasized that metadata helped them engage more deeply with medieval books. Working with original fragments, modern facsimiles, and digitized manuscripts, they came to argue that, in any medium, metadata is a book's voice—the tool by which that book tells users about itself, helping beginners understand and value what they see, helping experts uncover the complex web of connections that is the foundation for advanced research. While reading someone else's measurements of height and width may not help physically remote readers achieve what Treharne calls "optimal interpretative potential," recording this data shares important insights into those key details that cannot be discerned looking at pictures on screens.²⁵ Well before the rise of the so-called "digital age," print publications like J. J. G. Alexander's magisterial series *A Survey of Manuscripts Illuminated in the British Isles* gave manuscript scholars vitally important information, like the physical measurements, for distant manuscripts they also could neither see nor touch. In digital viewing environments, descriptive metadata fulfills a similarly important experiential function, helping bridge some of the inevitable gaps between eyes, mind, fingers, and screen.

Interoperability

"Interoperability" both in definition and in practice is as highly sought-after and elusive as a treasured relic in an Arthurian quest. The term's origins are rooted the mid-twentieth-century military discourse (in roughly the same decades that medievalists—valued for their practice "drawing conclusions from fragmentary evidence"—were

²² Edwards, "Back to the Real?" para. 6. The question of scale is further complicated by the fact that viewers may well use different sized screens to access the same digital copy of the same medieval book: smart phone, laptop, external monitor, tablet—all change the sense of how big the object might be and the viewer's bodily relation to it.

²³ Treharne, "Fleshing out the Text," 475–76.

²⁴ Treharne, "Fleshing out the Text," 476.

²⁵ Treharne, "Fleshing out the Text," 470.

entering the ranks of government programs like the Central Intelligence Agency).²⁶ Moving beyond Cold War military applications, the term was taken up in computing to describe “the ability of two or more computer systems or pieces of software to exchange and subsequently make use of data.”²⁷ In the context of information systems, the term now “denotes the ability of a system to work with or use parts of other systems.”²⁸ In digital libraries, “interoperability” has come to mean both the broader “potential for metadata to cross boundaries between different information contexts” and the much more specific ability “to exchange metadata between two or more systems without or with minimal loss of information and without any special effort on either system.”²⁹ Martin Foy defines “interoperability” as digital resources being “open in their data and connect[ing] to larger networks.”³⁰ To put it in more medievalist-friendly terms: consider the literary trope of the “envoy” in which the poet sends his or her creation to its intended audience, often by telling it, “go, little book.” *Interoperability* is the underlying philosophy—and hard work—that makes it possible for the (digital) book, once sent to a new home, to actually be read and understood with only minimal loss and transformation.

“Ther is so gret diversite in wrytyng of oure tonge,” Part I: Our Unalterable Inheritance of Codicological Variance

One of the largest challenges facing interoperability, and the creation of a single, universal digital manuscript hub, lies not in new technology but in old practice. For hundreds of years, the discipline of codicology has been characterized by widely divergent practices, with different schools and scholars developing extremely different ways of describing the same features in medieval manuscripts.

There is, for example, a centuries-long disagreement over what to call the animal-based membrane upon which a manuscript may be written. William Horman’s 1519 Latin-English phrasebook offers a range of terms one might choose to employ to describe this substance, all of which—significantly—he presents as equally valid:

That stouffe that we wrytte vpon: and is made of beestis skynnes: is somtyme called p[er]chement / somtyme velem / somtyme abortyue / somtyme

²⁶ On the military roots of “interoperability,” see John Unsworth, “Computational Work,” 1. See also, *OED* “interoperability” (n.1). On medievalists in the CIA, see Cantor, *Inventing the Middle Ages*, 261–62.

²⁷ *OED* “interoperability” n. 2.

²⁸ Haslhofer and Klas, “A Survey of Techniques,” 13.

²⁹ Haslhofer and Klas, “A Survey of Techniques,” 1, 8. See also Chan and Zeng, “Metadata Interoperability and Standardization” and Elings and Weibel, “Metadata for All.”

³⁰ Foy, 133. Foy briefly mentions metadata’s role in interoperability, but focuses his comments on new protocols like RDF and the creation of new metadata. How to continue to use the massive amount of already created metadata has yet to be raised in a sustained fashion by medievalists, even though the ability to continue to use important early digitization projects depends on engagement with legacy data as well as the latest database protocols.

me[m]bran. Parchement of the cyte: where it was first made. Velem / bycause it is made of a caluys skynne. Abortyue / bycause the beest was scante perfecte. Membraan / bycause it was pulled of [off] by hyl dynghe [skinning, or playing] fro the beestis lymmes.³¹

Some modern scholars share Horman's stance and insist that all these terms may be used interchangeably.³² Some use "parchment" as a generic term for any writing support made of animal membrane.³³ Others contend that "vellum" ought to be reserved for writing support made of calf skin, and use "parchment" to describe all other animal-skin writing support.³⁴ Still others draw a stricter distinction, reserving vellum for calf-skin and parchment only for goat or sheep.³⁵ In addition to disagreement between scholars, different libraries follow different cataloguing rules as they describe animal-skin writing supports.³⁶ As Christopher de Hamel notes, "In the manuscripts department of the Bodleian Library in Oxford the house usage today is to refer to the material consistently as parchment; in the British Library in London, the same substance is standardly called vellum."³⁷

Similar disagreements over practice occur in the proper methods for measuring leaves: millimetres, centimetres, or inches? The proper method for recording manuscript dates is also the subject of considerable variation. Ought palaeographers to use the traditional Latin abbreviations and roman numerals, like "s. XI"? Or Arabic numerals: "ca. 1250"? These different ways of recording manuscripts' dates vary between palaeographical schools. Sometimes, date usages also vary *within* the same palaeographical

31 Horman, *Vulgaria*, f.80v, via *Early English Books Online's* digital reproduction, of a microfilm reproduction, of the original print copy held in the Henry E. Huntington Library and Art Gallery.

32 See, for example, de Hamel, *Medieval Craftsmen*, 8.

33 See Clarkson, "Rediscovering Parchment," 5. The Library of Congress's *Thesaurus for Graphic Materials*, a key reference guide for metadata librarians, suggests "parchment" as the proper term for "skins and hides" and presents it as synonymous with, but preferable to, "vellum." Similar use of the term "parchment" for the entire genre of animal membrane can be found throughout da Rold, "Materials," and in Holsinger, "Of Pigs and Parchment."

34 See, for instance, the distinction drawn by Gillespie in "Bookbinding," 165n53.

35 On the terminology debates, see Ryder, "Parchment: Its History, Manufacture, and Composition," 392–93. Clemens and Graham also present an excellent concise overview of the terminology debates in *Introduction to Manuscript Studies*, 9–10.

36 Recently, shared standards for describing manuscripts have been created. However, these standards do not offer guidelines on the "parchment" versus "vellum" debate, preferring to leave that choice to local institutions. See *Descriptive Cataloging of Rare Materials (Manuscripts)*, 86. <http://rbms.info/files/dcrm/dcrmmss/DCRMMSS.pdf>. I am indebted to my colleague, and Binghamton University metadata librarian, Laura Evans for this reference. It is worth noting that although the standards do not differentiate between animal types for "parchment" or "vellum" or attempt to mandate what institutions ought to use, they appear to prefer "parchment" as their general term for animal-skin writing support. See, for example, the definitions of "membrane," "parchment," and "vellum," on pages 144, 145, and 147 respectively.

37 de Hamel, *Medieval Craftsmen*, 8.

school, even the same library, and the same book. For instance, both examples just cited—s. XI¹ and c. 1250—come from *A Catalogue of the Pre-1500 Western Manuscript Books at the Newberry Library*.³⁸

This variation can be problematic even when manuscripts are studied within their home institutions. It becomes even more problematic when these descriptions—as descriptive metadata—are brought together in a single digital manuscript aggregator. Aggregators like DMS-Index let researchers view side-by-side different manuscripts held at different institutions, and each of these manuscripts is described by its home institution's preferred codicological vocabulary. One knows, perhaps, at the Bodleian what “parchment” means, and at the British Library what “vellum” means—and, importantly, if one doesn't know one can ask a librarian. But what happens when users view these manuscripts from different institutions side-by-side in cyberspace, without the guiding wisdom of a human librarian well-versed in local cataloguing practice?

Picture a researcher uninterested in the long history of terminology variance in codicology. What this researcher is interested in, instead, is using the vast swathes of data now available online to analyze the regional differences in the use of calf and other animal skins in codex making. If she views side-by-side manuscripts from Harvard, the Huntington, the Bodleian, Cambridge University Library, the British Library, the Vatican Library, and the Bibliothèque Nationale de France, will she know what each of these libraries mean when they say “vellum” or “parchment”? Or, because she believes in the training she received at her particular palaeographical school, *and* because these institutions do not embed detailed explanations of their unique local practices within every single manuscript description they produce (because no one does), will she risk believing that “vellum” means something stable and dependable, across all institutions and manuscript she sees? Might she go on to trace animal-use variation based on the apparent difference highlighted in her manuscripts' metadata, when that variation, in fact, reveals nothing clear about the animals—but a great deal about curators and codicologists? And if she does go on to do this, who will be blamed when it is revealed that her data are irrevocably flawed? Certainly the researcher, but it seems quite likely that such research errors might also be used to single out digital manuscripts themselves as deceptive and flawed—when the problem lies much earlier, is in fact intertwined with and inseparable from the field.

Furthermore, more problems than just “parchment” versus “vellum” arise when combining manuscript descriptions made by different teams, at different moments, with different end goals in mind. The first iteration of DMS-Index successfully demonstrated interoperability by drawing together manuscript descriptions for medieval texts digitized by seven different partner institutions. But because of these partners' different local

³⁸ Example dates come from, respectively, the description for MS 1, a French book of homilies from the first quarter of the eleventh century, and the description for MS 19, a Franciscan Bible, also from France. Saenger, *A Catalogue of the Pre-1500 Western Manuscript Books at the Newberry Library*, 3, 35.

practices, this early DMS-Index retrieved records for 118 manuscripts on “parchment,” 14 more on “Parchment,” and 394 on “parch,” as well as 392 manuscripts on “Vellum” and 2 on “vellum”—rather than 920 manuscripts on the same animal-substrate³⁹ (if we allow ourselves to believe that “vellum” in this case means the same thing as “parchment,” which of course we should not). This wide range of search results occurred because computers cannot, without human help, intuit that “parchment” (lower-case p) is the same thing as “Parchment” (with an upper-case P)—let alone that both “parchment” and “Parchment” are the same thing as the convenient cataloguing shorthands “parch” and “perg.”⁴⁰ In their home environments, all of these terms are correct—ready, as it were, to go to the ball. But combined in an interoperable manuscript hub like DMS-Index, it is as though all of this metadata has, as John Unsworth puts it, “suddenly found itself in Union Station in its pajamas: it is not properly dressed for its new environment.”⁴¹

This data normalization issue is a familiar one to libraries and museums. To combat this almost gravitational pull toward variation, libraries and museums use “controlled vocabularies”: “an organized arrangement of words and phrases used to index content and/or to retrieve content through browsing or searching.”⁴² Professional cataloguers use these specific, authoritative, controlled vocabularies to make sure that their terminology fits their community’s accepted professional standards, which in turn allows their data to be findable across all other institutions that share the same vocabulary.⁴³ However, not all of the creators and manipulators of metadata share this training. (I did not, when I began my postdoc.) As digital humanists and other scholars cross into record creation and cataloguing, both the quality—and any deliberate or fetishized uniqueness—of the metadata we create can and will affect interoperability.

This variation in terminology can be addressed in a number of ways in digitization projects. First, scholars and digital humanists can use established controlled vocabularies in all new metadata we create. Second, all creators of new metadata—in digital projects, as well as in libraries and museums—can minimize future fragmenting of terminology by not resorting to shorthand like “parch” or “perg” for “parchment” (although the underlying problem of “parchment” and “vellum” will remain). Third, data curators can expect

39 This specific example of how terminological choices can limit broader interoperability comes from conversations with Albritton early in my postdoctoral work.

40 A similar issue appears in searching “language” across different institutions’ metadata. In the earlier iteration of *DMS*, researchers find 326 manuscripts by searching “Latin.” and 48 if one searches “Latin” without the final period. For the computers we humans partner with in our research, “Latin and English.” is not the same as “English and Latin.”—and neither are the same as “Latin and English” (without the final period), “Middle English with occasional Latin phrases,” “Latin, English.” or the machine-readable, iso-639-2 language codes “lat” and “eng.”

41 Unsworth, “Digital Humanities Centers as Cyberinfrastructure,” quoted in Unsworth, “Computational Work with Very Large Collections,” 4.

42 Harpring, *Introduction to Controlled Vocabularies*, 12.

43 Museums tend to rely on the *United List of Artists Names* and the *Thesaurus of Geographical Names*. Libraries turn to the *Library of Congress Subject Headings* and *Thesaurus for Graphic Materials*.

and uncover these (and other) data normalization problems and then build solutions into the code they write to transform manuscript descriptions for interoperable aggregators. Fourth, aggregators might move partner institutions' metadata through programs like "Open Refine," where manuscript records containing Parchment (with a capital "P") can be transformed *en masse* into "parchment" (with a lower-case "p") and "eng" and "lat" can become "English" and "Latin." Fifth, inter-institution projects can develop shared nomenclature, and projects that hope to share their manuscripts via aggregators like DMS-Index might take on the terminology already in use for the manuscript aggregator they hope to join (although that runs the risk of bending their immediate project *toward* future goals and away from *immediate* research needs and/or funders' requirements). Sixth, researchers can carve out part of their research time to look into terminology variation between the institutions they work with and make sure to do a version of data normalization in their own research. The list of potential fixes could likely be expanded almost infinitely.

Ultimately, however, as useful as these and other fixes will be for future digital manuscript work, the problem of human-introduced variation is likely to continue, rearing its head—Hydra-like—in new and interesting ways just when we believe we have solved it for good. This is in part because humans are prone to diversification, and humanists are especially fond of coining neologisms to fit the specific thing—or problem—we are grappling with in that moment. (Take, for example, the evolving debates over what one ought to call digital photographs of medieval books—are they digital manuscripts? Digital facsimiles? Digital surrogates? Avatars? Metaobjects?) Manuscript aggregators, thus, are a constant negotiation between the needs of humans, who as academics might build their careers on the coining of neologisms and new terminology, and the needs of machines, which require precision and repetition. Human researchers get far more from phrases like "Middle English with occasional Latin phrases,"⁴⁴ "Latin with some later scribbles in Middle English,"⁴⁵ "Latin with some English added on folios at the end,"⁴⁶ "Latin with Middle English verse in margin of F. 107v," "Latin and Old French with some English on the flyleaf under paste-down," than we do "Latin" and "English." But the computers on which we rely for searching can use the latter, and not the former.⁴⁷

Making a manuscript aggregator, thus, is an endless negotiation between the living and the dead, between human and machine, between different palaeographical camps and different cataloguers working in different decades, between digital humanists/scholars rewarded for uniqueness and information professionals and librarians

44 "Manuscript Description," Corpus Christi College, Parker Library, CCCC MS 282.

45 "Manuscript Description," Corpus Christi College, Parker Library, CCCC MS 21.

46 "Manuscript Description," Corpus Christi College, Parker Library, CCCC MS 112.

47 There is an important distinction between what is produced by a targeted search using the "Language" facet versus what the broader net of a keyword search might generate. A keyword search would catch manuscripts with language descriptions like "Latin and Old French with some English on the flyleaf under paste-down," while a targeted language-search would not. At the same time, this sort of broad keyword search also fails to distinguish between "Chaucer" as author/creator and "Chaucer" as subject, or "Chaucer" as largely irrelevant reference buried deep in the bibliography.

rewarded for regularity, between what we have inherited, and what we now want and need, and what we might guess our heirs in the future might want and need too. But that negotiation happens, strangely, off stage for most manuscript scholars. Despite the fact that we are trained to see these negotiations between past, present, and future as they play out in physical medieval books, we do not tend to seek them out as they pertain to new digital copies of old physical books. Despite our expertise in changes of older text technologies, these newer technological changes are invisible to us, even though our work is facilitated, and perhaps sometimes curtailed, by the invisible labour of data curation happening off screen.

One fairly simple intervention in the data problems produced by inevitable human variations is for all digitization projects to openly share all their underlying, messy metadata—as well as the cleaned-up, human-readable display seen in search screens.⁴⁸ But as important as that sharing is, it lays bare the problem. It does not fix it. Another, very different, intervention can be done by medievalists. Even as we seek to maximize data normalization in and between our digital manuscript projects, a medievalist perspective can also offer an important layer of practical resignation to the soaring ambitions of much digital work. As Boethius might note, only God is perfect and changeless, above the mutability of worldly things. The state of humans, by contrast, is constant change. Whether or not one agrees with Boethius about God, his point about the unavoidable changefulness of human work is useful to recall in digital projects. Remembering human mutability, or—to offer a different analogy, also connected to medievalists—accepting that there is no “one true ring to bind them all” and there never will be, can keep us focused on what we are actually doing and can do. We cannot revert to some state of prelapsarian, pre-Tower-of-Babel, never-existed-in-the-first-place grace where manuscript scholars always used the precise same terminology in precisely the same way. Instead, we must learn to respectfully reuse each other’s data as they are and not as we might idealistically wish them to be. To return to Boethius and Lady Philosophy’s final lesson: in medieval digital humanities we must always strive to do good, but we must understand that—in large and small ways—we will always fail.

“Ther is so gret diversite in wrytyng of oure tonge,” Part 2: The Unalterable Fact of Different Metadata Standards

There is another cause of variation that has to be understood when dealing with digital manuscripts. To search and discover digital manuscripts, object descriptions must be wrapped in XML tags that are shaped according to metadata standards that libraries, museums, and other cultural heritage institutions use to encode descriptive metadata. Strictly speaking, these metadata standards are *encoding* standards, which is to say, they are guidelines that determine what type of descriptive (and administrative and structural) metadata can be recorded in each file and where in each object description

⁴⁸ The Digital Walters and OPenn are important pioneers for this kind of open sharing of raw data.

this information must appear.⁴⁹ The best-known of these metadata standards include the Text Encoding Initiative (TEI), Dublin Core (DC), the Metadata Object Description Schema of the Library of Congress (MODS), Encoded Archival Description (EAD), and the Metadata Encoding and Transmission Standard (METS). This chapter focuses on the first three—TEI, DC, and MODS—because these are the standards I worked with closely in my time on DMS-Index. However, the work of metadata sharing across standards/schemas and my larger arguments are applicable to EAD, METS, and other forms of metadata and mark-up.

The foundational truth that humanities scholars must grasp about metadata is that even though there are established standards and controlled vocabularies and we would do well to follow them, there is no “one best metadata standard for all digital medieval projects”—and there never will be. As of 2010, there were “over a hundred community-recognized metadata standards from which to choose” and “additional metadata standards will continue to emerge as new classes of digital resources are created and as new user needs are identified.”⁵⁰ Therefore, this chapter vigorously refuses to endorse a single metadata standard for all digitized medieval manuscripts. While TEI may be the most familiar XML-metadata standard for digital humanists and has long played an important role in the marking up and sharing of medieval manuscripts, it is by no means the only way to render medieval manuscripts’ descriptive metadata.⁵¹ Museums, libraries, and archives have long had their own cataloguing methods, standards, and needs for medieval manuscripts. As all these libraries, digital humanities projects, and cultural heritage institutions have moved into manuscript digitization, they have carried their cataloguing practices and metadata schemas with them.

Perhaps the second most important thing for medievalists and digital humanists to understand about metadata standards is that each of these standards was developed at a specific moment in history, in response to a particular community’s concerns and specific pressures, in order to serve that community’s larger needs and goals. Furthermore, as communities and needs have evolved, so too have the standards. Understanding—and sharing—digital manuscripts requires historicizing digitization projects *not* in the time that the manuscripts were created, but within the specific historical moment of the digital project’s creation. That moment, that community, and those goals all determine not just the

49 In other words, encoding standards will not tell you *how* that metadata ought to be encoded. TEI and DC do not tell you what to call parchment or whether you can abbreviate to “parch” or “perg” without inhibiting interoperability. They simply provide you with a field or a tag where you can enter the information as you (ideally with recourse to your field’s controlled vocabularies) see fit.

50 Cole and Han, *XML for Catalogers and Metadata Librarians*, 96. Digital Librarian Jenn Riley has created a remarkable visualization of the 105 most-discussed and best-known metadata standards used by cultural heritage institutions for different objects and different communities: “Seeing Standards: A Visualization of the Metadata Universe.”

51 The most recent major revision of the TEI guidelines includes a chapter dedicated to the best practices, as defined by the TEI consortium, for describing medieval manuscripts. P5, first released November 2007; at time of writing most recently updated March 3, 2016.

method of digitization but the form that the metadata takes. And the form that metadata takes determines how shareable and discoverable these digital resources really are.

Writing the full history of each schema/community is beyond the scope of this chapter. However, some basic knowledge of who these communities were (and still are), as well as the problems each metadata standard exists to help fix, reveals how impractical, impossible, and frankly undesirable it would be to attempt to select one universal metadata standard. Rather than engaging in a quixotic quest to erase the inevitable variations of individual practice, a better solution to interoperability begins with understanding the diverse communities of practice that have come together in shared digital manuscript environments, because this confluence of communities, goals, metadata standards, and values continues to shape—visibly and invisibly—the digital objects that increasingly drive our research.

TEI

Perhaps the most well-known of these XML-based metadata standards within digital humanities and literary studies is the Text Encoding Initiative (TEI). TEI is an internationally recognized set of guidelines developed for encoding electronic texts, such as novels, plays, and poetry: it functions primarily in support of humanities research.⁵² TEI was not created specifically for medieval manuscripts, nor for encoding their physical descriptions—although the fifth iteration of the TEI guidelines includes an extensive chapter dedicated to methods for encoding the physical descriptions of medieval manuscripts.⁵³ Since TEI was not created for medieval books, projects that use TEI for medieval manuscript metadata adapt techniques for encoding literary content to encode bibliographic content.

TEI was established in 1987, in response to the rapid growth of a wide variety of digital technologies “dominated by mutually incompatible formats.”⁵⁴ These incompatible formats imposed “serious technical obstacles even to the simple transfer of data files from one machine to another, to say nothing of the difficulties posed by mutually incompatible and proprietary file formats.”⁵⁵ Against this bubbling sea of one-off programs, TEI was developed to aid mutual comprehension, information sharing, and the building of tools that could be used on more than one humanities computing project. TEI’s founders came from a variety of academic backgrounds.⁵⁶ Their ambitious goal was to create “practical recommendations as to how an extensible set of guidelines consistent with the goal of a universal text-encoding scheme might be achieved.”⁵⁷ Such ambition is very much in keeping with late 1980s computing, in and beyond humanities disciplines.

52 NISO, “Understanding Metadata,” 4.

53 TEI: Text Encoding Initiative. “[Chapter 10: Manuscript Description](#).”

54 Burnard, “The Evolution of the Text Encoding Initiative,” 3.

55 Burnard, “The Evolution of the Text Encoding Initiative,” 3.

56 Burnard, “The Evolution of the Text Encoding Initiative,” 2–3.

57 Burnard, “The Evolution of the Text Encoding Initiative,” 3.

As visionary as the creators of TEI were, they were not the only people in the late 1980s with “an expertise in the creation and management of digital text” concerned by the ways that a widespread lack of standardization was inhibiting the universal sharing of text and data.⁵⁸ Nor were they the only community seeking to aid mutual intelligibility and information sharing through metadata standardization across projects and institutions.

DC

Twenty years earlier, in 1967, the presidents and library directors of colleges and universities in Ohio created a nonprofit dedicated to “develop[ing] a computerized system in which the libraries of Ohio academic institutions could share resources and reduce costs.”⁵⁹ Dubbed the Ohio College Library Center, its shared cataloguing system debuted in 1971 and revolutionized library cataloguing and academic research, shepherding in (among other things) Interlibrary Loan as we know it today.⁶⁰ By the early 1990s, the Ohio College Library Center had expanded to all fifty states and had a new name: the Online Computer Library Center (OCLC), which is perhaps best known today as the organization behind the WorldCat global library catalogue.⁶¹ It is important to realize that the information professionals of the OCLC were not unaware of the standardization that troubled the founders of the TEI, but—for their community—that was not the real problem. Well before the digital revolutions that convulsed the 1980s, they had already developed a standard system for data sharing.

The problem that librarians faced, instead, was the creation and growth of the World Wide Web. By 1994 approximately half a million addressable objects were online, but they were not yet mutually comprehensible and discoverable.⁶² To solve this problem, in 1995 the OCLC and the National Center for Supercomputing Applications (NCSA) held a joint workshop in Dublin, Ohio.⁶³ Participants there developed a concise “standard set of 15 interoperable metadata elements designed to facilitate the description and recovery of document-like resources in a *networked* environment,” which they

⁵⁸ Burnard, “The Evolution of the Text Encoding Initiative,” 3.

⁵⁹ Rosenheck, “OCOC: From an Historical Perspective,” 2.

⁶⁰ This revolutionized the role of computers in academic libraries in two important ways: first, it “enabled libraries to rapidly and efficiently catalog books” because each library no longer had to catalogue by hand every book they received. If the book was already in the shared catalogue, they could use existing cataloging information. If it was not in the shared catalogue, they could catalogue it so that other partner libraries would be able to save time by reusing that information. Second, it “provided location information for the materials listed in the catalog by participating libraries:” users were suddenly able to easily find and request materials held at another partner library. Rosenheck, “OCLC: From an Historical Perspective,” 2.

⁶¹ For views on how OCLC and WorldCat transformed twentieth-century librarianship, see *What the OCLC Online Union Catalog Means to Me*.

⁶² “History of the Dublin Core Metadata Initiative,” Dublin Core Metadata Initiative.

⁶³ “History of the Dublin Core Metadata Initiative,” Dublin Core Metadata Initiative.

named Dublin Core (DC) after their meeting place.⁶⁴ Today, DC is “by far the most used descriptive metadata standard in the library domain,”⁶⁵ and in 2012 expanded to include a total of fifty-five elements and element refinements, in order to provide more, and more specific, metadata. Again, this metadata standard was not invented with medieval manuscripts in mind. But, because—when digitized and put online—manuscripts are, in fact, “document-like resources in a networked environment,” DC is an important player in digital manuscript metadata. This importance is also due to the fact that the Open Archives Initiatives Protocol for Metadata Harvesting (OAI-PMH), “which supports metadata sharing and interoperability,” “mandates the use of Simple Dublin Core as its lingua franca, or lowest-common-denominator XML metadata grammar.”⁶⁶ In other words, if a library or institution believes in the broader mission of the Open Archives Initiative, and wants all content in their repository to be interoperable with the enormous network of repositories that are part of the OAI, they must do at least part of their metadata according to the constraints of Simple DC.

Despite its deep roots in data sharing between academic libraries, DC is not *the* default metadata standard for all academic libraries, let alone all libraries, museums, and cultural heritage institutions that hold, digitize, and display content online. This is, in part, because within library communities “there has historically been some tension between supporters of a minimalist view” of metadata creation, “who emphasize the need to keep the elements to a minimum and the semantics and syntax simple,” and “supporters of a structuralist view, who argue for finer semantic distinctions and more extensibility for particular communities.”⁶⁷ The same concision that makes DC such a useful tool for sharing basic metadata across institutions limits the complexity of descriptions fit into DC. What this means is that DC, either in its original “Simple” 15-element form or in its expanded “Qualified” 55-element form, is an excellent solution for some kinds of metadata problems and goals, but not others.

MODS

In 2002, the Library of Congress developed the Metadata Object Description Schema (MODS) as a middle path between the forceful simplicity of DC and the complexities that many cataloguers and metadata librarians still desired.⁶⁸ As the National Information Standards Organization puts it,

Rich description of electronic resources is a particular focus of MODS, which provides some advantages over other metadata schemes. MODS elements are

⁶⁴ “Dublin Core (DC),” *Dictionary for Library and Information Science*, 234, my italics. See also, “Understanding Metadata,” 3.

⁶⁵ Following the descriptive metadata system known as MARC. Cole and Han, *XML for Catalogers and Metadata Librarians*, 98.

⁶⁶ Cole and Han, *XML for Catalogers and Metadata Librarians*, 101.

⁶⁷ NISO, “Understanding Metadata,” 3.

⁶⁸ The history and use of MARCXML, while not part of this chapter, is covered extensively in Cole and Han.

richer than the Dublin Core; its elements are more compatible with library data than ... Dublin Core standards; and it is simpler to apply than the full MARC 21 bibliographic format.⁶⁹

MODS, that is, helped update and streamline earlier, complicated bibliographic formats while simultaneously giving libraries a richer set of elements and attributes for describing their holdings than the fifteen elements of Simplified DC.⁷⁰ Once again, MODS was not created with medieval manuscripts in mind, but digitized manuscripts are, in fact, “electronic resources,” and it makes sense for the creators and maintainers of library catalogues who use MODS to make their digital manuscript holdings discoverable alongside their other physical and electronic holdings.

If TEI was created to fit the needs of humanists and academic researchers and DC was created to solve the problems of information professionals seeking to harness the networking power of the internet, MODS was created to serve the needs of individual libraries, which need to be able to catalogue diverse resources throughout their collections in ways that facilitate searching while not flattening out the complexities of vastly different objects. Like TEI and DC, MODS was not created specifically for describing medieval manuscripts, but as digitized manuscripts enter library catalogues—if those catalogues are already filled with thousands, or millions, of items described according to the local interpretation of MODS—then digitized manuscripts’ metadata will likely be slotted into MODS elements, so that patrons can find these digital resources when they search the library’s catalogue.

To reframe these brief histories: it bears repeating that *none* of these standards were developed for medieval manuscripts, nor medievalists, nor manuscript scholars. MODS was developed by librarians and cataloguers for librarians and cataloguers. It offers more opportunity for granularity and fine-grained details than DC, but that granularity comes at a cost: even as best practices and rules for implementers proliferate, “MODS records from different institutions and different projects tend to be less consistent at this time.”⁷¹ By contrast, DC was also developed by librarians, in partnership with information technologists in academia, government, and industry. DC creates opportunities for impressive regularity across wildly diverse digital collections held at different institutions and repositories. This cross-collection regularity, however, comes at the cost of the granularity in individual description. TEI, by contrast, was developed largely outside of academic libraries—it is primarily by (humanities) researchers for (humanities) researchers. Its flexibility makes it endlessly adaptable to particular projects’ and researchers’ needs, but often has little to do with libraries’ and/or cataloguers’ needs.

69 NISO, “Understanding Metadata,” 5–6.

70 In the increasingly complicated timeline of evolving metadata standards, this places the creation of MODS before the expansion of DC to “Qualified Dublin Core,” with its fifty-five elements. Qualified DC was created after MODS, and fulfills some, but not all, of the same granular functionality of MODS.

71 Cole and Han, *XML for Catalogers and Metadata Librarians*, 111. For the MODS best practices document I consulted, see the Library of Congress, “DLC/Aquifer Summary of MODS Requirements and Recommendations Table.”

As data curator for DMS-Index, my job was to aggregate metadata from different collections, marked according to different metadata standards, which were always already shaped by very different communities, goals, and needs (see, for example, [Table 8.1](#)). The Walters Art Museum marked up its descriptive metadata according to TEI P5 in order “to ensure broad access to [their] manuscript descriptions” and ensure “that [their] manuscript descriptions be available to a wide range of users and can be rendered in multiple output formats (such as PDFs, HTML web pages, and page turning applications).”⁷² However, because Stanford Library’s Digital Object Registry represents its descriptive metadata in MODS, and because Stanford and the Walters entered into a data sharing agreement shortly before my postdoc began which allowed Walters manuscripts to be discoverable through Stanford’s catalogue, I needed to turn the Walters Art Museum’s TEI into Stanford’s local implementation of MODS. Because Stanford Libraries was also home to the first iteration of DMS-Index, there were good institutional reasons for continuing to have MODS be the metadata standard underlying DMS-Index. Parker on the Web, by contrast, was significantly older than the other digital manuscript repositories this chapter covers. It predated the TEI P5 guidelines used by Walters and e-codices, and—adhering to accepted TEI standards when the Parker Project began—encoded existing manuscript catalogues in the same way other digital humanities projects at the same time were encoding novels, plays, and poems. Because ecodices is an inter-library digital repository, its needs as a collaboration of more than forty libraries across Switzerland made DC a logical metadata choice. At the same time, a profound need for highly detailed manuscript descriptions inspired the additional use of TEI P5 as a second, more granular metadata standard. Reading these repositories’ medieval manuscript metadata in this way, with and against each other, and with profound interest in the logic behind each repository’s choice of schema, demonstrates, as Johanna Drucker has eloquently argued, that

metadata schemes must be read as models of knowledge, as discursive instruments that bring the object of their inquiry into being, shaping the fields in which they operate ... Analysis of metadata and content models, then, is an essential part of the critical apparatus of digital humanities.⁷³

Crosswalks and Transforms

Because there never has been, and never will be, a single metadata standard appropriate to the needs of all projects, objects, and stakeholders, libraries have developed terminologies and best practices for taking information expressed in one metadata standard and remaking that same information according to the rules and constraints of another metadata standard. The first step is creating “a crosswalk.” A crosswalk, as defined by the National Information Standards Organization (NISO), is

a mapping of the elements, semantics, and syntax from one metadata scheme to those of another. A crosswalk allows metadata created by one community

72 “Describing Manuscripts with TEI,” The Digital Walters.

73 Drucker, *SpecLab*, 11.

to be used by another group that employs a different metadata standard. The degree to which these crosswalks are successful at the individual record level depends on the similarity of the two schemes, the granularity of the elements in the target scheme compared to that of the source, and the compatibility of the content rules used to fill the elements of each scheme.⁷⁴

Also sometimes called “field mapping” or “metadata mapping,” a crosswalk is generally represented as a table, or a chart, showing “equivalent or nearly equivalent metadata elements or groups of metadata elements within different metadata schemas.”⁷⁵ The phrase—“nearly equivalent metadata elements”—might seem to suggest a close, or close-enough, fit. But in reality, different metadata standards can present the same information in markedly different ways (see [Table 8.1](#), row 3).

Moreover, crosswalk building is rarely as simple as these basic descriptions might suggest: metadata is created at different times, by different institutions, according to different standards, to serve different needs. Thus, different sets of metadata rarely fit together seamlessly or, in some cases, without a great deal of shoving. A particular challenge for medieval manuscript aggregators is the fact that, as NISO notes “the mapping of schemes with fewer elements (less granularity) to those with more elements (more granularity) is problematic.”⁷⁶ For example, the DC element <creator> is “an entity primarily responsible for making the resource.”⁷⁷ In a single-author illustrated manuscript, <creator> encodes not just author, but also translator(s), artist(s), and/or scribe(s) since they are also, according to long-standing custom in manuscript studies, also acknowledged as “entities primarily responsible for making” the manuscript. In manuscripts that compile works by different authors, or that were copied by several different scribes, or that contain work by more than one artist, the list of “entities primarily responsible for making” tagged by <creator> explodes. By contrast, in TEI’s most recent guidelines for manuscript descriptions there is a much more targeted element for the specific creator-figure <author>. Since authors are not the only humans responsible for the making of medieval books, institutions like the Walters Art Museum have developed ways of creatively adapting the TEI element <respStmt> to acknowledge the unique contributions of manuscripts’ scribes, translators, and artists. E-codices uses a different TEI element—<persName role= “author”>—to achieve the same precision. And, of course, all of these contributors are acknowledged in a different way in MODS (see [Table 8.2](#)).

The crosswalk, or mapping, is only first step for sharing metadata and manuscripts between institutions. After a crosswalk has been written showing how the data ought to be transformed, the data still has to actually *be* transformed. In my work on DMS-Index, this was done by writing a custom XQuery script (also called a “transform”) in the program oXygen for each institution’s digital manuscript collection. For each institution I created one script, which would take tens, hundreds, sometimes more than

⁷⁴ NISO, “Understanding Metadata,” 11.

⁷⁵ See “crosswalk” and “metadata mapping” in “Glossary,” Gilliland, *Introduction to Metadata*.

⁷⁶ NISO, “Understanding Metadata,” 11.

⁷⁷ “Term name: creator,” dublincore.org.

Table 8.1. Comparison of XML-encoded metadata schemas employed by participating institutions, along with a sample of how each institution represents the concept/field of the maker of the intellectual work in their local interpretation of their preferred metadata schema(s).

Institution	Stanford University	Walters Art Museum	Parker on the Web	e-codices
XML Metadata Schema	MODS	TEI P5	TEI, P4	OAI:DC TEI P5
Sample crosswalk	<name> <namePart>Boethius </namePart> <role> <roleTerm>author </roleTerm> </role> </name>	<author>Boethius </author>	<author>Boethius </author>	<creator>\Boethius<creator> AND ALSO <author>Boethius</author>

Table 8.2. Sample crosswalking of how the three example participating institutions mark up a work’s creator(s), according to each institution’s local interpretations of their chosen metadata schema(s).

Metadata Standard	Walters TEI (P5)	Stanford MODS	e-codices Simplified DC	e-codices TEI, P5
Crosswalk for the concept “author”	<author>	<name type=“personal”> <namePart> <role> <roleTerm>author	<creator>	<author> AND ALSO <persName role=“author”>
Crosswalk for the concept “scribe”	<respStmt> <resp>scribe	<name type=“personal”> <namePart> <role> <roleTerm>scribe	<creator>	<persName role=“scribe”>
Crosswalk for the concept “artist”	<respStmt> <resp>artist	<name type=“personal”> <namePart> <role> <roleTerm>artist	<creator>	<persName type=“scribe”>
Crosswalk for the concept “translator”	<respStmt> <resp>translator	<name type=“personal”> <namePart> <role> <roleTerm>translator	<creator>	<persName role=“translator”>

1,000 individual metadata documents, and make new versions of those descriptions in new metadata schema. This step is known as “transforming” the metadata. For those concerned about digital *mouvance* run amok, it is important to note that this process does not fundamentally change the parent record. It is creating a new copy of the manuscript description, in the new schema, that—ideally—loses or changes as little of the original content as possible. However, as will become clear in my discussion of transforming the Walters Art Museum digital manuscript metadata, some data loss is almost always inevitable in metadata transformations.

Sympathy for Adam Scriveyn: Transforming the Walters Metadata

As a medievalist specializing in English literature, I work in a field haunted by lost books: from the Reformation reconfigurations and destruction of medieval libraries, to the Cotton fire of 1731, to the mid-century socialist biblioclasm of Otto Ege and the ongoing, present-day commercial destruction of books like the Hungerford Hours.⁷⁸ And, despite my straightforward pronouncements above on inevitability of loss and change, when I began curating manuscript metadata for DMS-Index I felt the weight of centuries of destruction bearing down on me. Although I knew, intellectually, that the loss of descriptive metadata crosswalked into a new digital environment was *not* actual physical destruction enacted on medieval books, I still felt I had to preserve all that was known about each digital medieval manuscript assigned to my care. If I failed, it seemed a betrayal of my professional, and ethical, duty to serve and preserve the past and the dead.⁷⁹

My preservation principles were not merely driven by dread of professional failure. They were also influenced by what can only be described as a kinship I was beginning to feel with medieval scribes—particularly those cursed by medieval authors and modern editors for their failures to properly copy the text. I knew that, at least in some circles, my failures in metadata curation might be cursed as quickly and smoothly as Chaucer (if in fact he wrote the lyric) curses his scribe in “Adam Scriveyn.”⁸⁰ The final line of the short poem describes the scribe’s poor copying as the result of the Middle English

78 On the Renaissance reconfiguration of medieval libraries, see Summit, *Memory’s Library*. On the Cotton fire, see Prescott, “Their Present Miserable State of Cremation.” On Ege, see Gwara, *Otto Ege’s Manuscripts*. See also Benjamin Albritton’s work digitally reconstructing Ege’s dismembered books and Morcos, “Piecing Together the Puzzle of the Hungerford Hours.”

79 Kiss et al. touch on similar dilemmas and fears of data loss faced in textual editing as it moves from print to digital modes of production. See “Old Light on New Media,” 20. Searle’s “Possible History” was a touchstone text, discussing medievalists’ duty to the past and the dead. See especially Searle, “Possible History,” 779.

80 By and large, the fifteenth-century scribe John Shirley’s attribution of the poem to Chaucer has been accepted in modern scholarship. For a provocative case against attributing the poem to Chaucer, see Edwards, “Chaucer and ‘Adam Scriveyn,’” 135–39.

word “rape,” which can mean either “haste” or “sexual assault and violence.” In my first weeks transforming manuscript metadata, I dreaded being indicted for enacting similar violations on texts given to my care.

Beyond lingering fears of being cursed by my colleagues, my developing curatorial principles were also influenced by my research in fifteenth-century English poetry and poetics, in which there is a strong precedent for expansive principles of preservation. Writing on John Lydgate, David Lawton offers backhanded, yet entirely accurate, praise: “this is one of his most impressive characteristics, that he is no trimmer.”⁸¹ Although praise of Lydgate’s expansive principles of inclusion tends to be shadowed with irony, my engagement with Lydgate’s curatorial practice was decidedly straightforward. As a translator and transmitter of inherited continental literary traditions, Lydgate was no trimmer; as a data curator, it turns out, neither was I. Thus, with Adam Scriveyn as my negative exemplar and John Lydgate as my positive model, I sought to reproduce the entirety of the original metadata entrusted to my care.

At times, when the different elements used by the Walters Art Museum’s particular dialect of TEI and those favoured by Stanford’s local interpretation of MODS matched closely, it was easy to adhere to this idealistic vision of data curator-as-perfect-scribe. For instance, the Walters’ TEI element `//msContent/msItem` may be crosswalked directly to the Stanford MODS element `//relatedItem type=“constituent.”`⁸² Although the match is not quite as seamless between the ways the different metadata standards denote authors (see Table 8.2, above), it is still straightforward and fairly easy to carry over that information from the Walters TEI to Stanford MODS.

However, as the granularity and richness of the TEI metadata increases, the ability of MODS to contain that data lessens. The problem is, from the standpoint of the crosswalk builder, that there are no matching elements in MODS for TEI’s specialized elements for describing writing support, dimensions for the average leaf size and the written space, foliation, collation, signatures, catchwords, layout, hand description, and decorations. Nor does MODS support the kind of nesting of items within items within items that TEI allows. Because it was created to serve broader collection needs than just those of medieval manuscripts, MODS does not have pre-existing elements or attributes to be pulled into service here. Instead, recognized MODS elements had to be bent and stretched into scarcely recognizable shapes to fit my exacting curatorial principles. My commitment to aggressive preservation paired with the formal constraints of MODS, fostered within my Walters TEI-Stanford MODS crosswalks an increasingly long and baggy collection of `<note>` elements, all modified by the attribute `@displayLabel`, in order to contain all possible manuscript descriptors (see Table 8.3).

Readers might be tempted here to dismiss MODS as a useful tool for manuscript metadata, but that would miss my point. A rigorous codicology of the digital manuscript

⁸¹ Lawton, “Dullness and the Fifteenth Century,” 779.

⁸² Here, for simplicity’s sake, I show the XPath for the particular example lines, as they appear in the Walters and Stanford’s XML metadata documents. This is not how it is actually encoded in the XML itself in either institution.

Table 8.3. Sample crosswalk of Walters TEI to Stanford MODS demonstrating how heavily the @displayLabel attribute was worked to hold important metadata that had no other place in the MODS.

Walters TEI	Stanford MODS
<msContents> <textLang>	<note displayLabel="Language Note">
<msContents> <msItem> <locus>	<note displayLabel="Locus">
AND <msContents> <msItem> <msItem> <locus>	
<msContents> <msItem> <rubric>	<note displayLabel="Rubric">
AND <msContents> <msItem> <msItem> <rubric>	
<msContents> <msItem> <incipit>	<note displayLabel="Incipit">
AND <msContents> <msItem> <msItem> <incipit>	
<physDesc> <objectDesc> <supportDesc> <extent>	<note displayLabel="Foliation Statement">
<physDesc> <objectDesc> <supportDesc> <extent> <dimensions type="leaves"> <height unit="cm"> <width unit="cm">	<note displayLabel="Dimensions">

(continued)

Table 8.3 (Cont.)

Walters TEI	Stanford MODS
<code><physDesc></code> <code><objectDesc></code> <code><supportDesc></code> <code><extent></code> <code><dimensions type="written"></code> <code><height unit="cm"></code> <code><width unit="cm"></code>	<code><note displayLabel="Writing Block"></code>
<code><physDesc></code> <code><objectDesc></code> <code><supportDesc></code> <code><collation></code> <code><p></code> <code><formula></code>	<code><note displayLabel="Collation: Formula"></code>
<code><physDesc></code> <code><objectDesc></code> <code><supportDesc></code> <code><collation></code> <code><p></code>	<code><note displayLabel="Collation: Notes"></code>
<code><physDesc></code> <code><objectDesc></code> <code><supportDesc></code> <code><signatures></code>	<code><note displayLabel="Signatures"></code>
<code><physDesc></code> <code><objectDesc></code> <code><supportDesc></code> <code><collation></code> <code><p></code> <code><catchwords></code>	<code><note displayLabel="Catchwords"></code>
<code><physDesc></code> <code><objectDesc></code> <code><layoutDesc></code> <code><layout @columns></code>	<code><note displayLabel="Layout: Columns"></code>
<code><physDesc></code> <code><objectDesc></code> <code><layoutDesc></code> <code><layout @ruledLines></code>	<code><note displayLabel="Layout: Ruled Lines"></code>

(continued)

Table 8.3 (Cont.)

Walters TEI	Stanford MODS
<code><physDesc></code> <code><objectDesc></code> <code><layoutDesc></code> <code><layout></code>	<code><note displayLabel="Layout: Notes"></code>
<code><physDesc></code> <code><bindingDesc></code> <code><binding></code>	<code><note displayLabel="Binding"></code>
<code><physDesc></code> <code><decoDesc></code> <code><decoNote></code>	<code><note displayLabel="Decoration Note"></code>
<code><history></code> <code><origin @notBefore @notAfter></code> <code><origDate></code>	<code><note displayLabel="Date Note"></code>
<code><history></code> <code><provenance></code>	<code><note displayLabel="Provenance"></code>
<code><history></code> <code><acquisition></code>	<code><note displayLabel="Acquisition"></code>

does not retrospectively judge *whether* MODS was “the right choice” (whatever that might mean), but rather seeks to understand *why* MODS was chosen. What larger social and institutional pressures made MODS the correct choice for these project-builders in this moment? And what might that choice reveal about the larger structures in place that support the creation of digital manuscripts? And, perhaps most pressing, what does this metadata standard reveal about the nameless, unseen humans behind the digital manuscript we see hovering on our screens?

One notable discovery, once one digs into the underlying philosophies of information organization revealed by an institution or project’s chosen metadata schema, is that TEI manuscript-description may ultimately suggest collections of like-objects grouped with like. Nineteenth-century novels read against other nineteenth-century novels. Medieval manuscripts grouped with other manuscripts. These TEI descriptions are rich and lovely, but may also be somewhat cut off from the larger transhistorical movements of text and text technologies by virtue of that same delicious precision. By contrast, because MODS is for describing libraries’ collections—across time and media—use of this metadata schema for medieval manuscripts highlights similarities between medieval manuscripts and other, non-medieval, non-manuscript items. TEI celebrates manuscripts’ difference from all other forms of text. MODS highlights the similarities. The choice of one schema tunnels users ever deeper into the minutiae of medieval books, the other moves users along a much broader and more diverse continuum of text. Both have their gifts, as well as their costs. But rather than making a swift comparison and then dismissing MODS as

the wrong tool for the wrong job, a codicology of digital medieval books along the lines of what I advocate considers all of these things.

Even with my heavy reliance on MODS `<note @displayLabel>` to contain the information tagged by the Walters in separate TEI elements, the full granularity and precision of the Walters' metadata could not always be preserved. A particular challenge lay in the `<note>` elements nested inside `<msItem>` elements (which appeared alongside other nested elements like `<incipit>`, `<rubric>`, and so on). The Walters cataloguers and metadata creators drew distinctions between different types of notes: `<note type="content">`, `<note type="text">`, and `<note type="work">`.⁸³ MODS has a `<note>` element, which I was already using enthusiastically for things like `<note displayLabel="rubric">`, so in theory I could easily turn the Walters' ever-more granular `<note>` elements into matching MODS `<note>` elements. However, if I were to transfer into MODS the Walters' careful distinctions between `@type="text"` and `@type="content"`, I would need to create separate `<note>` elements for each type: the Walters `<note type="content">` *could* be matched to a Stanford `<note displayLabel="Note(s) about Manuscript's Content">`, the Walters `<note type="text">` to a Stanford `<note displayLabel="Notes about Textual Variation">`. But should it? While this would preserve the Walters' elegant precision, at a certain point I risked capsizing the metadata's usefulness under the weight of an expanding series of notes about notes about notes about notes.

If this project was merely about one institution's data, such a solution might be feasible. However, because my early work was laying a foundation for a much larger interoperable environment, I could not design with just the Walters in mind. In my work with the Walters, I was also designing what might be considered the general "metadata envelopes" that I would go on to reuse for all subsequent institutions who shared their manuscripts and metadata via DMS-Index. While I included a statement encouraging users of DMS-Index to seek out further information with the original holding institution, the fact remained that the majority of readers in DMS would not know, at a glance, what the Walters' distinctions between "text," "content," and "work" meant when read alongside manuscripts from Oxford, Cambridge, and the Bibliothèque Nationale de France, or any other institutions who might not draw the same distinctions in their metadata.⁸⁴

Consider the profoundly distorting effect on readers if they were viewing a Walters manuscript alongside one from the Lund University Library in Sweden. The Walters uses meticulous sub-types of notes within notes to highlight different information. To achieve a similar effect, the Lund Library uses a single `<note>` element with each `<msItem>`, and

83 See The Digital Walters, "Describing Manuscripts with TEI."

84 The obvious response and solution here is project documentation, but this is not always enough. The Walters is a model in meticulous documentation. This documentation stands outside the individual metadata documents I was processing. Therefore, because it functions as a kind of digital paratext, standing outside the text of the metadata proper, this documentation would not automatically be seen in DMS. While I included links in each transformed metadata set back to the original metadata record, these links only mean that readers of Walters metadata in DMS would have the opportunity to follow links back to the original metadata in its original environment and from there might dig out the original project documentation.

within that <note> parent element employs the child element <p> to indicate a break. Both institutions have similar goals, but use different XML tagging solutions to achieve them. But the formal properties of the XML are not commensurate. Imagine if I preserved all the Walters' separate sub-types of <note> and simply matched them to the Lund <note>. An unintended effect of my structural preservation of the Walters' original form might be that the Lund's cataloguing would look less meticulous and rigorous in comparison to the Walters because the Walters had many categories to the Lund's one overarching information unit. Alternately, the effect might also be that the Walters could appear curiously fragmented, and therefore lessened, against the Lund's unitary vision of what the <note> should be and do. Neither case would be true: my metadata forms would be participating in a kind of suggestive untruth, pushing readers to see difference rather than showcasing coherence across continents, countries, and libraries. Faced with the need to create metadata that would serve across standards and across different institutions' cataloguing and XML approaches, I created a single MODS <note displayLabel="Catalog Note(s)"> to contain all the Walters' TEI <note> elements. I did so to create a standard metadata form that would serve all institutions who chose to share their manuscripts and metadata. But I also knew that in so doing I was failing to create a perfect copy of the metadata entrusted to my care. To serve the greater project, my zeal for perfect changeless preservation had to begin to crack.

Similar challenges emerged when dealing with metadata concerning the Walters' manuscripts' languages. The Walters systematically seeks to serve the needs of human researchers and computer systems by pairing machine-readable information with human-friendly nuanced text.

`<textLang otherLangs="dut" xml:id="n496.203230" mainLang="lat">The primary language in this manuscript is Latin. The secondary language of this manuscript is Dutch; Flemish.</textLang>`⁸⁵

It is simple to write an XQuery script crosswalking the information enclosed by the TEI element <textLang> tag into similar tags using DMS-Index-style MODS. However, machine-comprehensible information about manuscripts' languages was locked within the @otherLangs and @mainLang attributes inside the Walters <textLang> element. The easiest way to move this information into MODS would have been to command oXygen to directly copy the three-letter language codes ("dut" and "lat") from the @mainLang and @otherLangs attributes. But this would serve neither the needs of humans, nor computers. After all, computers do not know that "lat" means "Latin" unless humans tell them. So, "the easy fix" would result in "lat" and "dut" being listed as languages, instead of "Latin" and "Dutch" or "Flemish," thus repeating and exacerbating all the problems of "parchment" versus "parch" described earlier.

Even more than being a good, dutiful scribe and "no trimmer," my goal as data curator was to begin to help undo part of the digital Tower of Babel that develops when different institutions' metadata jostles together in interoperable settings like DMS. Therefore,

85 W.169, my italics.

I wrote a series of *if-then-else* statements in my XQuery script that ran through every possible three-letter language code in the Walters digitized manuscripts, in both @mainLang and @otherLangs attributes. (Doing so added more than 100 lines to my script.) Each time my program saw “lat,” “dut,” “gre,” or any other three letter language code in the inherited, “parent” metadata, its job was not to copy that language code into MODS element but to write in a new line of metadata. When the code saw “lat” in Walters TEI, it wrote in Stanford MODS

```
<language>

      <languageTerm>Latin</languageTerm>

</language>
```

When it saw “arm” in Walters TEI, it wrote

```
<language>

      <languageTerm>Armenian</languageTerm>

</language>
```

and so on. In short, to preserve the metadata, I had to change the metadata. *Mouvance* was inevitable. Despite my deep desire to dutifully copy my exemplars, even when those exemplars were characterized by beautifully organized XPaths, interoperability mandated that I could not *not* be Adam Scriveyn.

Curation is Not the Same as Hoarding: Lessons Learned from Parker on the Web

Unlike the Walters, the descriptive metadata for manuscripts accessible via Parker on the Web is not produced according to TEI P5, [Chapter 10](#) guidelines. This is because the Parker on the Web app predates the creation of those TEI P5 guidelines on how to describe medieval manuscripts. Instead, the Parker team drew on what was considered best practice at the time the project was being designed, using then-current TEI P4 guidelines to encode M. R. James’s print catalogue (1909–1912), with some changes, additions, and customization.⁸⁶ What this means is that the formal structures of the

⁸⁶ “Manuscript Descriptions,” in “About the Project: Project Technology,” Parker Library On the Web: About the Projects. The decision has been the subject of pointed critique, most notably in Ken Pennington and A. S. G. Edwards’ 2013 *TLS* exchange (Edwards’ piece can be found at: www.the-tls.co.uk/articles/private/better-on-the-continent/. Pennington’s response is at: www.the-tls.co.uk/articles/private/lets-get-digital-2/. Edwards’ reply is at: www.the-tls.co.uk/articles/private/digital-texts/.) For a reasoned consideration of the social and economic pressures that influenced this decision by the Parker team, see Rundle, “Virtual Manuscripts and the Real World. Part I.”

Parker library metadata are very different from those of the Walters Art Museum's metadata. While the Walters presents its metadata as short phrases nested along regular XPaths, and these XPaths are consistent across manuscript descriptions, Parker on the Web marks up prose. In the Parker on the Web metadata, long paragraphs of description are punctuated with TEI elements bracketing important search information.⁸⁷

Due to the funding structures and project mission for my position, I wrote only one transform per repository, which meant that my transforms were limited to copying data that occurred in predictable XPaths across the majority of manuscripts in that repository. For instance, James's manuscript descriptions often include a heading for Decoration Notes, which, when it is used, appears consistently along the XPath

```
/TEI/text/body/msDesc/p[@function="Decoration"]/decoNote
```

Using this XPath, my XQuery transform could pick out all information found between the <decoNote> tags at this path and then copy that same information into the appropriate MODS elements using the following command:

```
{for $f in $y/text/body/msDesc/p[@function="Decoration"]
return
<note displayLabel="Decoration Note">{data($f/fn:normalize-space())}</note>}
```

which, essentially, tells oXygen to take whatever information it finds along the XPath `/TEI/text/body/msDesc/p[@function="Decoration"]` and copy it into the space between the start-tag `<note displayLabel="Decoration Note">` and the end-tag `</note>`. However, a number of manuscripts in Parker on the Web do not have the element `<p function="Decoration">`: from them, the command would collect no information along this path. Some of these manuscripts without the element `<p function="Decoration">` contain decoration notes elsewhere in their metadata. For example, Corpus Christi Collect Cambridge, Parker Library MS 153 has no `<p function="Decoration">` but has two `<decoNote>` elements retrievable at the XPath `/TEI/text/body/msDesc/p/decoNote`.⁸⁸ Decoration notes for Corpus Christi College Cambridge, Parker Library, MS 304 appear not at `/TEI/text/body/msDesc/p[@function="Decoration"]` but instead at `/TEI/text/body/msDesc/p[@function="Research"]/q/decoNote`. For the data curator tasked with writing a single transform to work for an entire collection, such variation in XPaths is a significant challenge, because it means that similar information is stored at very different places, all of which need to be taken into account.

87 Paratextual elements, including paragraph breaks and running headnotes of the pages of the James catalog, are also all preserved in Parker's XML tags.

88 See lines 96 and 98 in the descriptive metadata.

Furthermore, taking these other paths “into account” is not just a matter of obsessively close reading metadata patterns until all possible options emerge.⁸⁹ In the Parker on the Web metadata, some of the information framed by granular tags like the <decoNote> mentioned above would be of extremely limited—or no—value once it has been extracted from its original setting and served out in an interoperable manuscript hub. For example, the descriptive metadata for Corpus Christi College Cambridge, Matthew Parker Library MS 304, as it is marked up in the Parker on the Web’s original P4 TEI, has eight extractable <decoNote> elements along the XPath /TEI/text/body/msDesc/p[@function=“Research”]/q/decoNote. But while the first is of undeniable value for researchers encountering the manuscripts’ metadata in an external manuscript hub like DMS-Index, the latter seven contain content of limited value for reuse outside Parker on the Web (see [Table 8.4](#)).

Again, I do not want to be misunderstood. This is not a critique of what the Parker team did or did not do in marking up their metadata. Although those critiques have been, and continue to be, levelled (and may be of some value for new projects seeking models), such critiques are not particularly useful for anyone seeking to use the metadata of the Parker Project manuscripts *as it exists* today. One may dream of an ideal world in which every digitization and digital humanities project meticulously designs its metadata to simultaneously serve the needs of the project at hand, as well as all possible needs of all possible future inheritors of that data. One may also dream of an ideal world in which legacy projects, like the Parker, are given so many financial resources that they can hire a host of workers to retype the entire Parker Library catalogue according to the current TEI P5 manuscript description guidelines. While one is dreaming, one might also dream up a related endowment so that the Parker, and all libraries around the world like it who were early innovators in digitization, can be rewritten every time metadata standards evolve or the TEI changes. In the world in which we actually live and work, however, projects come into being in their own historical moment; their design is driven by that project’s particular needs and goals; and their metadata is shaped accordingly.

As a data curator, and as a literary scholar interested in the long evolution of text technologies, I have no interest in critiquing this project or arguing for that new method or standardization. Instead, I am interested in the concrete facts that underscore all digital work, and which therefore underscore the possibilities of that universal portal to all manuscripts that medievalists speak of in hushed and excited tones. Not all data can be reused. The painful truth is that, for various reasons, there will always be some data that cannot be translated from one project to another. And, in fact, in the case of <decoNote> here, translating that metadata into the new schema would result in a less useful final product for researchers using an aggregator like DMS-Index, a manuscript description riddled with enigmatic fragments that served former project needs but make little sense in the new system. As I worked with the metadata of the Parker Library, I set aside my motto of “be no trimmer” and—in collaboration with my project manager Benjamin

⁸⁹ Taken to the extreme and committing to doing this to all 500+ manuscripts currently available in Parker on the Web, this approach would be extremely time-consuming, and still would not guarantee perfect extraction of the metadata.

Table 8.4. Descriptive metadata on the decorations and embellishments of Corpus Christi College Cambridge, Parker Library MS 304 as that information appears in the Parker on the Web TEI compared to what information would appear once this metadata had been extracted and transformed to DMS-Index MODS. Emphasis in TEI column added.

Decoration Notes as they appear in Parker on the Web TEI (surrounding context included)	Extractable information as it would appear in DMS-Index MODS
<decoNote> <i>The three preliminary leaves are ornamentally written in large capitals (11 lines to a page), within borders, the first and last pages being enclosed in an arch, the rest within rectangular borders</i> </decoNote>.	<note displayLabel="Decoration Note"> The three preliminary leaves are ornamentally written in large capitals (11 lines to a page), within borders, the first and last pages being enclosed in an arch, the rest within rectangular borders </note>
on the opposite page the text of <name type="author">Iuuencus</name> begins without any rubric, the first preface being written <decoNote> <i>in the same fancy capitals</i> </decoNote> as the verses of <name type="person">St Isidore</name>	<note displayLabel="Decoration Note">in the same fancy capitals</note>
<locus>4v</locus> <name type="person">Iohannis</name> fremit ñ misteria uitae caluetii: aquilini: : siue iuueni: so far <decoNote> <i>in large fancy capitals</i> </decoNote>, then immediately in ordinary uncials: Immortale nihil ñ	<note displayLabel="Decoration Note">in large fancy capitals</note>
then at once in larger capitals: sapientissimi uiri Iuueni: xpiani: euangeliorum liber primus: explicit Incipit liber secundus caluetii ~ aquilini ~ Inde philippus ait >>> these last words are <decoNote> <i>in red</i> </decoNote> but in ordinary uncials	<note displayLabel="Decoration Note">in red</note>
the 2nd book ends, and the 3rd begins thus (67<abbr rend="superscript">a</abbr><locus>67r</locus>): turbasque reliquit. Explt Incpt liber tertius (<decoNote> <i>in red</i> </decoNote>)	<note displayLabel="Decoration Note">in red</note>
the 3rd book ends, and the 4th book begins (93<abbr rend="superscript">b</abbr><locus>93v</locus>): hominum seletio fiet: Euangeliorum liber tertius explct (<decoNote> <i>in red</i> </decoNote>)	<note displayLabel="Decoration Note">in red</note>
Incpt eiusdem liber ...(<decoNote> <i>in black</i> </decoNote>)	<note displayLabel="Decoration Note">in black</note>
quartus? feliciter? (<decoNote> <i>in red</i> </decoNote>). Talia ñ The MS. breaks off book IV, verse 733	<note displayLabel="Decoration Note">in red</note>

Albritton—developed a new motto: “curation is not the same as hoarding.” Everything cannot be kept. Users of digital manuscripts and their metadata would do well to recall Ariel’s song from *The Tempest*:

Full fathom five thy father lies;
Of his bones are coral made;
Those are pearls that were his eyes:
Nothing of him that doth fade,
But doth suffer a sea-change
Into something rich and strange.⁹⁰

Creating new copies, physical or digital, is always a process of loss and change.

However, while some loss is inevitable, there are data that cannot be lost: for instance, manuscripts’ dates of creation. M. R. James dates his manuscripts in charmingly idiosyncratic phrases—and in roman numerals. In Parker on the Web these date notes are tagged in XML, using the same method as the <decoNote> tagging, which helps human readers but does not help a computer tasked with finding manuscripts from a certain century. By this time in my term as data curator, I was working with XQuery functions, which allowed me to help machine applications focus on standardized data within idiosyncratic human phrasing. In the Parker transform, I had hoped to use fn:contains to pull the date’s data from within the shifting human words and recopy it into machine-friendly form. I found, however, that the underlying logic of roman numerals stood in the way of my proposed solution. Since “xiii” contains “xii,” “xi,” and “x,” the new digital humanities skills I had been honing did not help me with this part of the Parker metadata (although they were exceedingly useful elsewhere). In projects like these, when one works on grant money for the set term of two years, time spent learning new skills to pursue the particular challenges of one institution’s metadata set is useful to a point: beyond that, it is a field of diminishing returns. In time-sensitive, grant-funded work, excessive time expended on one data set is time that cannot be given to other libraries’ metadata. In the case of the Parker on the Web, rather than continue to spend more time and grant money learning new techniques (say, a reverse case statement, which would have helped here) in hopes that they *might* work, after a certain period of time I moved forward with the skills I had. With Albritton’s support, I ceased trying to pull dates from the XML marked-up James’s catalogue and instead pulled the Parker manuscripts’ dates from the Parker on the Web’s .html (which had the added benefit of containing revised dates based on new research and better information than what James used).

Significantly, our solution echoes medieval copying practices of combining flawed or fragmentary exemplars to create a better copy. For example, midway through Thomas Hoccleve’s *Series* (early 1420s), the narrator Thomas shows a character known as “the Friend” the new English translation of a story from the *Gesta Romanorum* that the Friend asked Thomas to create.⁹¹ Thomas tells the Friend that the poem is complete. However, the Friend objects that Thomas’s new translation lacks the moralization that the Friend treasures in his personal copy of the *Gesta*. The Friend lends this second copy

⁹⁰ Shakespeare, *The Tempest*, I.ii. 474–79.

⁹¹ See “Fabula de Quadam Imperatrice Romana” in *My Complaynte and Other Poems*, lines 953–80.

to Thomas, whereupon Thomas copies the missing portion into his book. From two Latin exemplars arises Thomas's single English text. Neither Thomas nor the Friend are troubled by Thomas's failure to make a perfect copy of either exemplar. Precise copying of one parent manuscript is not the point for them. Combining the best parts of both to create a newer and more useful copy is.

Similarly, in drawing dates from the Parker .html and everything else from the Parker's TEI-marked-up James's catalogue, I found myself making a "better" object by copying parts of two separate documents. There are perhaps no downsides here. The dates are more correct. The manuscripts more discoverable. But in taking the best data from two different copies I had to let go of what I had seen as a scribe-like duty to make a *precise* copy in order to make a more *useful* copy. Over the months of my postdoc work, both my skills as a copyist and my understanding of the kind of copying I was supposed to be doing were evolving. I no longer saw data curation as a duty to strictly preserve inherited documents. Instead, my duty was to create documents that could be *used*.

To put it less like a data curator and more like a medievalist, in working with the Parker on the Web metadata for DMS-Index, I moved past the association of copyist-induced change with betrayal, loss, and violence that I had drawn from the Chaucerian lyric "Adam Scriveyn." Instead, my models became Chaucer's heirs: Thomas Hoccleve and John Lydgate. In the prologue to *The Fall of Princes*, Lydgate (translating and transforming Laurent de Premierfait, who in turn is translating and transforming Boccaccio) describes the positive reasons that copyists change old books as they make them anew:

Artificeres hauyng exercise
May chaunge and turne bi good discrecioun
Shappis, formys, and newli hem deuyse,
Make and vnmake in many sondry wyse,
As potteres, which to that craft entende,
Breke and renewe ther vesselis to a-mende.

Thus men off crafft may off due riht,
That been inuentiff & han experience,
Fantasien in ther inward siht
Deuises newe thoruh ther excellence;
Expert maistres han therto licence
Fro good to bettir for to chaunge a thyng,
And semblabli these clerkis in writyng,

Thyng that was maad of auctours hem beforn,
Thei may off newe fynde and fantasie,
Out of old chaff trie out ful cleene corn,
Make it more fressh and lusti to the eie,
Ther subtil witt and ther labour applie,
With ther colours agreable off hewe,
Make olde thynges for to seeme newe.

[Craftsmen, having practical experience,
May change and turn by good judgment

Shapes, forms, and newly construct them
 Make and undo in many different ways—
 Like potters, who are inclined to that craft,
 Break and revise their vessels to emend.

Thus men of craft may of due right
 (That are inventive and have experience)
 May think up in their inward sight
 New designs through their excellence.
 Experienced masters have the right
 From good to better to change a thing—
 And similarly, these clerks in writing

A thing that was made by earlier authors/authorities
 The clerks may, through new discoveries and invention,
 Out of old chaff produce very pure corn,
 Make it more vivid and pleasing to the eye
 Their clever judgment/understanding and their labor apply
 With their colours of pleasing hue
 Make old things appear new.⁹²

Like Lydgate's "artificeres," "men off crafft," and "clerkis in writyng" who "make and unmake in many sondry wyse," curating manuscript metadata for reuse in interoperable environments requires—and likely will always require—a certain amount of breaking, renewing, and amending. Rather than see that breaking and renewal as a betrayal of the medieval past, I propose that closer attention to the hidden labour of metadata transformation underlying digital manuscripts reveals just how deeply digital medieval book culture is rooted in copying practices akin to those practiced by medieval writers, scribes, and bookmakers.

The Tower of Babel in the Catalogue: e-codices and the Place of English in Digital Manuscript Studies

E-codices, the digital medieval manuscript library of Switzerland, collects the digitized manuscripts of more than fifty-eight libraries and private collections, both in and outside of Switzerland. A basic level of manuscript metadata is shared via the Open Archive Initiative Dublin Core (OAI:DC) metadata standard: manuscript's title, short description/abstract, author or creator, publisher, date, and format, which includes material support and foliation statement. This DC-tagged information is in English, and is lean and accessible across institutions. However, because it has been carefully shaped to fit the goals and requirements of OAI:DC, this first layer of metadata does not match the scope of the descriptive metadata DMS-Index contained drawn from the Walters Art Museum and Parker on the Web. For instance, e-codices' top-layer DC metadata does not contain a full statement of contents, detailed decoration descriptions, or detailed

92 Lydgate, *Fall of Princes*, bk 1, lines 9–28.

Table 8.5. Mock-up of what that macaronic description, drawing together English-language OAI:DC and French-language TEI, might look like.

<abstract> This ethical work by Boccaccio, originally written between 1353 and 1356 and expanded in 1373, addresses the subject of the unevenness of fate. Manuscript copies of the work were frequently made; it was issued in print and translated into many languages. It enjoyed great popularity in Europe. The French translation by Laurent de Premierfait for Jean de Berry was equally popular, as evidenced by the 68 manuscript copies of this text still in existence....</abstract>

...

<note displayLabel="Catalog Note(s)">Traduction par Laurent de Premierfait exécutée en 1409. Ce manuscrit ne comporte pas le premier prologue de Laurent de Premierfait (dédicace de sa traduction).</note>

...

<note displayLabel="Foliation: Statement">Foliotation récente au crayon de 10 en 10. Les deux feuillets de garde, au début et à la fin, sont contemporains du manuscrit. Le feuillet 338v° est blanc.</note>

...

<note displayLabel="Hand Note">Ecriture bâtarde. Une seule main.</note>

provenance information. This information, instead, appears in a second set of metadata documents: extended, granular descriptions in TEI P5 rich enough to empower the far-reaching research that interoperable hubs like DMS-Index seek to support. Thus, by pairing two different metadata standards, to support two different kinds of interaction with their data, e-codices manages to serve the inter-repository needs of simplicity, and individual researchers' needs for scholarly depth. Using the same approach I had developed on the Parker Library metadata, I was determined to pull the best information from the DC and the TEI, seeking to make the richest possible metadata for future researchers using DMS-Index.

However, dipping into these richer TEI manuscript descriptions brought a new curation challenge. As the e-codices personnel with whom I corresponded had warned, these extended TEI descriptions were often not in English, but in German, Italian, and French. For example, Cologne, Fondation Martin Bodmer, Cod. Bodmer 174, a fifteenth-century manuscript of Laurent de Premierfait's French translation of Giovanni Boccaccio's *Des Casibus Virorum Illustrium*, has a short description in English tagged in the OAI:DC metadata schema. It also has a more extended manuscript description in TEI, which contains more granular information akin to what I transformed from the Walters Art Museum and Parker on the Web. This extended TEI manuscript description is in French. Although I was technically able to combine the <abstract> from the English-language OAI:DC with the cataloguing and content details supplied in the longer French-language TEI description, the resulting MODS metadata would have been macaronic (see [Table 8.5](#)).

For most data curators, this kind of linguistic mash-up would be immediately and obviously untenable. Yet I wondered if it could be defended. Medieval writers wrote macaronic verse. Some of the most famous and influential "English" medieval manuscripts

are, in fact, multilingual, and the recovery of medieval multilingualism is an important focus in the field.⁹³

Ultimately, however, DMS-Index was not just for experts who have the languages to tackle medieval and modern multilingualism. In helping build the next iteration of DMS-Index, I was hoping to create something also useful for undergraduates, book artists, calligraphers, and interested hobbyists. But seeking this broader audience imposed strict limits on what could be defined as “usable” metadata. What would students in a medieval literature course or an introductory book history course do when set to research using digital sources? The formal language of manuscript description already carries a high cognitive load for beginners.⁹⁴ Asking new recruits to medieval studies to also engage with formal manuscript descriptions that slide from English into Italian, German, and French would likely push that cognitive load to the breaking point. Accessibility does not just mean putting manuscripts on the web. It means surrounding those manuscripts with information that acts as a doorway in for new users. Words, as Geoffrey of Vinsauf puts it in his *Poetria Nova* (1208–1213), should work to render difficult ideas accessible.⁹⁵ His early thirteenth-century counsel on the ornaments of style offers an important critique of language used to render medieval manuscripts “accessible” in the digital age:

If, therefore, you introduce any words that are strange or recondite, you are displaying your own virtuosity thereby.... Take counsel: it may be you know all things—you are greater in others than this—still, in your mode of expression be one of those others.... Regard not your own capacities, therefore, but rather his with whom you are speaking.⁹⁶

Seeking to open up manuscripts to a wider range of possible users, I abandoned the macaronic manuscript catalogue.

This decision is not ideologically neutral. There are important, growing critiques of digital humanities that highlight its problems of monolingualism, monoculturalism, and the dominance of English as its *lingua franca*.⁹⁷ In seeking to better serve more diverse

93 See, for example, Butterfield, *The Familiar Enemy*.

94 For this point, in particular, I am indebted to the students in my “Medieval Books in the Digital Age” course, Binghamton University, Fall 2015. While I have taught manuscripts at other universities and to other students who I am certain also found manuscript descriptions difficult to parse, these students were uniquely honest about their difficulties parsing the stylized language of formal manuscript description. My thinking on language, access, and medieval studies is also informed by Carruthers, “‘Micrological Aggregates’: Is the New Chaucer Society Speaking in Tongues?”

95 Geoffrey of Vinsauf, *Poetria Nova*, 54–55.

96 Geoffrey of Vinsauf, *Poetria Nova*, 55.

97 Exploration of the ways that pervasiveness of English and Anglo-American interests deform digital humanities threads throughout Fiormonte, Numerico, and Tomasi, *The Digital Humanist*; see especially “Conclusions—DH in a Global Perspective,” 207–18. See also Fiormonte, “Toward a Cultural Critique of Digital Humanities,” 59–76; as well as his blog posts “Digital Humanities in Kerala: Some Lessons from the South” and “Towards a Monocultural (Digital) Humanities?” and Columbia’s blog post “Postcolonial Studies, Digital Humanities, and the Politics of Language.”

audiences within the English-speaking world, my curation decisions also reified in DMS-Index—at least for now—the problematic state of English as a default language of international scholarship.⁹⁸ Put differently, in seeking to solve one problem of the lack of diversity in the academy, my choices contributed to another.

Conclusion

I remain profoundly displeased by the choices that I made regarding the e-codices metadata. In the same way, I am haunted by other curation decisions I made as data curator for digital metadata from many other institutions generously sharing their manuscripts with DMS-Index. There are good reasons why I made these decisions, and in the previous pages I hope I have suggested some of those:

- The ever-present panting of “time’s winged chariot” that one lives under on a short-term project supported by external grant money and how that pressure influences project forms and timelines
- The slow-growing DH- and data curation skills that beginning digital humanists can only learn on the job through the process of trial, error, and *making*, and the impact that this learning will inevitably have on project processes and final productions
- The idea that we do not digitize these manuscripts solely for scholars, and my metadata is meaningless if it does not help recruit future generations to the study and care of medieval literature and culture.

But these good reasons do not erase the fact that I have equally good reasons to continue to be haunted.

In the end, I would argue that there is great value in publicly owning the haunting choices we make on digital projects. It is *good* for digital humanists to be haunted by the choices we make. Exorcism is not—should not—be the point: instead, we need to continue to discuss those hauntings and ethical compromises publicly. It is wrong to hide these ghosts because, in the case of my work on DMS-Index and other labour on digital manuscripts, hiding our ghosts contributes to an incorrect impression that all digital manuscripts can be all things, to all people, in the first iteration. As a corrective, those of us involved in digital projects and digitization need to carve out spaces to openly reflect on the compromises we have made, under what conditions, and for what reasons, while perhaps also laying out ways that we hope future iterations and future makers might improve on our compromises and alleviate our betrayals. When information is excised and done so consistently, that needs to be made transparent in multiple areas. It is of paramount importance to be as clear—as honest—about any modifications or transformations as possible.

By explicitly using medieval writers like Chaucer, Hoccleve, and Lydgate to theorize digitization and manuscript metadata, I hope I have also shown two additional

⁹⁸ See, for instance, Bocanegra-Valle, “English is my default academic language,” and Clavert, “The Digital Humanities Multicultural Revolution Did Not Happen Yet.”

things: first, how much of one's own world view and research digital makers bring to their projects. Just as data is not neutral, making digital things is never neutral. DH projects are shaped by who we are, and what we value and need, as we engage in the act of building. After seven years' intensive study on medieval literature and culture, I could no more stop being a medievalist as I did my digital work than I could stop breathing—and (even if those medievalist principles are not, or at least not yet, explicitly described alongside my work in DMS-Index) my deep reading in late medieval English literature acted as a lodestar for my nascent theories of digital book making.

Second, by reflecting on how much Chaucer, Hoccleve, and Lydgate shaped my curatorial work, I hope to have demonstrated more broadly just how much the cultures of pre-modern book-making have to offer contemporary digital humanities. The rhetoric of DH and digitization can slip into technological determinism and techno-utopianism. Against these impulses, medieval writers' insights into the labour, purpose, and limitations of book-making can serve as a useful corrective to modern visions of unchecked progress. Furthermore, when medieval works are marked up for further processing using modern computer languages and technologies, editors are engaging in a centuries-old practice that they may not be entirely aware of. Developing a richer, more self-conscious awareness of the long history of those practices can in turn inform and temper some of the central presumptions about digital technologies today—including any claim that the present is a break with the past simply because today's work happens to be digital. Bringing this prehistory into contemporary conversations on digitization helps shift modern discussions away from wide claims of democratization to more nuanced consideration: *democratizing for whom? To what end? At what cost—both literally, in terms of dollars or pounds expended to achieve project goals, and figuratively. Who, or what, has been treated as expendable?*

Moreover, making those inevitable compromises explicit does not just allow modern users to become more literate in digital manuscripts: it embeds the digital manuscript more overtly, and more firmly, in the historical moment of its own making, revealing the pressures under which all digital books are produced and circulated. The exploration of networks of making and circulation is, of course, a central credo of book historians. By making both our data and statements on the compromises that have shaped them openly available—participating in a self-conscious and rigorous codicology of digital medieval books—medievalists can do important work demystifying what has, too often over the last few decades, been treated as a transparent or simple process. Medieval manuscripts on screens are, fundamentally, different objects from the physical ones they seek to represent. As we leave the incunabula age of digitization, the time has come to not just *say* this—but to treat them as such, and to study their creation for insights into medieval labour and bookmaking, and into the invisible making that increasingly shapes our scholarship today.

Finally, there is an important ethical aspect to fostering a richer and more rigorous curiosity into the digital labour and labourers that make and maintain digital medieval books. Digitizers and data curators are inescapably part of the present history of the digitized manuscripts upon which much of our research depends. They are also inescapably part of our world today. The photographers who helped digitize the Matthew Parker

Library at Corpus Christi College, Cambridge are Polish.⁹⁹ At the time of writing, some of them are still working as digitizers at Cambridge University Library. The “Brexit” vote of June 2016 raises questions about their future employment. It also raises questions about their safety, as anti-Polish and anti-immigrant violence in Britain surges in the wake of that referendum. As I write these lines in New York in November 2016, the United States is facing its own wave of post-election, racist, anti-immigrant, anti-Muslim, and anti-woman violence. How many medievalists and manuscript scholars know whether American-based digitizers of medieval books will find themselves targeted in the weeks and months to come? (How many of us even know these digitizers’ names?) Whether or not we pre-modern book historians and digital humanists choose to make these connections, we are implicated in the infrastructures and politics that shape the lives of these modern makers of medieval books. We dedicate our lives to studying all that can be discovered about the processes of bookmaking and the makers of medieval books more than half a millennium ago: we should be able to extend that care to the modern copyists of medieval books today.

In this call to action, I have no medieval antecedent. Medieval book history is teeming with unnamed and uncredited labourers. Alexandra Gillespie has noted that the individual agency generally attributed to single individuals—especially printers, such as “Caxton,” “de Worde,” “Grafton,” “Pynson,” etc.—is a convenient shorthand for much more complicated relationships among networks of labourers.¹⁰⁰ As Gillespie insists, whenever these names are used they “should be understood as collaborative (involving other publishers, other printers, correctors, compositors, foremen, apprentices).”¹⁰¹ Unfortunately, by-and-large they are not—Caxton is lionized as a unique genius and the workers who actually made the type and pressed the books fade into shadows from which they cannot return. But in this next push to copy medieval books into new media, it need not remain this way. We have a chance, in this next chapter of medieval book history, to rectify some of these long-standing wrongs and fill in those vexing silences with living voices. Ultimately, this is my challenge for medievalists, manuscript scholars, and digital humanists: that we expand our circle of inquiry to include living bookmakers as well as the dead, in order to write them—to invite these bookmakers to write themselves—back into the long history of medieval books.

Bibliography

- Bailey, Moya Z. “All the Digital Humanists Are White, All the Nerds Are Men, but Some of Us Are Brave.” *Journal of Digital Humanities* 1 (2011). <http://journalofdigitalhumanities.org/1-1/all-the-digital-humanists-are-white-all-the-nerds-are-men-but-some-of-us-are-brave-by-moya-z-bailey/>.
- Bocanegra-Valle, Ana. “‘English is my Default Academic Language’: Voices from LSP Scholars Publishing in a Multilingual Journal.” *Journal of English for Academic Purposes* 13 (2014). 65–77.

⁹⁹ Personal conversation, Suzanne Paul, Cambridge University Librarian, October 20, 2016.

¹⁰⁰ Gillespie, *Print Culture and the Medieval Author*, 25–26.

¹⁰¹ Gillespie, *Print Culture and the Medieval Author*, 25–26.

- Bredehoft, Thomas. *The Visible Text: Textual Production and Reproduction from Beowulf to Maus*. Oxford Textual Perspectives. Oxford: Oxford University Press, 2014.
- Burnard, Lou. "The Evolution of the Text Encoding Initiative: From Research Project to Research Infrastructure." *Journal of the Text Encoding Initiative* 5 (2013). <https://jtei.revues.org/811>.
- Butterfield, Ardis. *The Familiar Enemy: Chaucer, Language, and Nation in the Hundred Years War*. Oxford: Oxford University Press, 2010.
- Cantor, Norman. *Inventing the Middle Ages: The Lives, Works, and Ideas of the Great Medievalists of the Twentieth Century*. New York: Harper Perennial, 1991.
- Carruthers, Mary. "'Micrological Aggregates': Is the New Chaucer Society Speaking in Tongues?" *Studies in the Ages of Chaucer* 21 (1999). 1–26.
- Chan, Lois Mai, and Marcia Lei Zeng. "Metadata Interoperability and Standardization: A Study of Methodology, Part I: Achieving Interoperability at the Schema Level." *D-Lib Magazine* 12 (2006). www.dlib.org/dlib/june06/chan/06chan.html.
- Chaucer, Geoffrey. *Troilus and Criseyde*. In *The Riverside Chaucer*. Edited by Larry D. Bensen. 3rd ed. Boston: Houghton Mifflin Company, 1987.
- Clarkson, Christopher. "Rediscovering Parchment: The Nature of the Beast." *The Paper Conservator* 16 (1992). 5–26.
- Clavert, Frédéric. "The Digital Humanities Multicultural Revolution Did Not Happen Yet." *L'histoire contemporaine à l'ère numérique* (blog), April 26, 2013, <http://histnum.hypotheses.org/1546>.
- Clemens, Ray, and Timothy Graham. *Introduction to Manuscript Studies*. Ithaca: Cornell University Press, 2007.
- Cole, Timothy W., and Myung-Ja K. Han. *XML for Catalogers and Metadata Librarians*. Third Millennium Cataloging Series. Santa Barbara: Libraries Unlimited, 2013.
- Corpus Christi College, "Manuscript Description: Language." CCCC MS 2I. Parker on the Web. https://parker.stanford.edu/parker/actions/manuscript_description_long_display.do?ms_no=2I.
- . "Manuscript Description: Language." CCCC MS 282. Parker on the Web. https://parker.stanford.edu/parker/actions/manuscript_description_long_display.do?ms_no=282.
- . Parker Library. "Manuscript Description: Language." CCCC MS 112. Parker on the Web. https://parker.stanford.edu/parker/actions/manuscript_description_long_display.do?ms_no=112.
- da Rold, Orietta. "Materials." In *The Production of Books in England 1350–1500*, edited by Alexandra Gillespie and Daniel Wakelin, 12–33. Cambridge: Cambridge University Press, 2011.
- de Hamel, Christopher. *Medieval Craftsmen: Scribes and Illuminators*. Toronto: University of Toronto Press, 1992.
- Descriptive Cataloging of Rare Materials (Manuscripts)*. Chicago: Rare Books and Manuscripts Section of the Association of College and Research Libraries, 2016.
- Dictionary for Library and Information Science*, edited by Joan M. Reitz. "Dublin Core (DC)," 234–35. Westport, CN: Libraries Unlimited, 2004.
- Drucker, Johanna. *SpecLab: Digital Aesthetics and Projects in Speculative Computing*. Chicago: University of Chicago Press, 2009.

- Dublin Core Metadata Initiative. "History of the Dublin Core Metadata Initiative." <http://dublincore.org/about/history/>.
- Echard, Siân. "Coda: The Ghost in the Machine: Digital Avatars of Medieval Manuscripts." In *Printing the Middle Ages*, 198–216. Philadelphia: University of Pennsylvania Press, 2008.
- . "Containing the Book: The Institutional Afterlives of Medieval Manuscripts." In *The Medieval Manuscript Book: Cultural Approaches*, edited by Michael Johnston and Michael Van Dussen, 96–118. Cambridge: Cambridge University Press, 2015.
- Edwards, A. S. G. "Back to the Real?" *Times Literary Supplement*, June 7, 2013, www.the-tls.co.uk/articles/public/back-to-the-real/.
- . "Chaucer and 'Adam Scriveyn.'" *Medium Aevum* 81 (2012): 135–39.
- Elings, Mary, and Gunter Weibel. "Metadata for All: Descriptive Standards and Metadata Sharing across Libraries, Archives and Museums." *First Monday* 12 (2007). <http://firstmonday.org/article/view/1628/1543>.
- Fiormonte, Domenico. "Digital Humanities in Kerala. Some lessons from the South." Infolet (blog), November 25, 2015. <https://infolet.it/2015/11/25/digital-humanities-in-kerala-some-lessons-from-the-south/>.
- . "Toward a Cultural Critique of Digital Humanities." In *Controversies around the Digital Humanities*, edited by M. Thaller. Special issue, *Historical Social Research/Historische Sozialforschung* 37 (2012): 59–76.
- . "Towards a Monocultural (Digital) Humanities?" Infolet (blog), July 12, 2015. <https://infolet.it/2015/07/12/monocultural-humanities/>.
- Fiormonte, Domenico, Teresa Numerico, and Francesca Tomasi. *The Digital Humanist: A Critical Inquiry*. Translated by Desmond Schmidt and Christopher Ferguson. Brooklyn: punctum books, 2015.
- Foys, Martin. "Media Archaeology and the Digital Incunable." In *The Medieval Manuscript Book: Cultural Approaches*, edited by Michael Johnston and Michael Van Dussen, 119–39. Cambridge: Cambridge University Press, 2015.
- Geoffrey of Vinsauf. *Poetria Nova*. Translated by Margaret F. Nims. Toronto: Pontifical Institute of Medieval Studies, 1967.
- Gillespie, Alexandra. "Bookbinding." In *The Production of Books in England 1350–1500*, edited by Alexandra Gillespie and Daniel Wakelin, 150–72. Cambridge: Cambridge University Press, 2011.
- . *Print Culture and the Medieval Author: Chaucer, Lydgate, and Their Books, 1473–1557*. Oxford: Oxford University Press, 2006.
- Gilliland, Anne J. "Setting the Stage." In *Introduction to Metadata*, edited by Murtha Baca, 1–19. 2nd ed. Los Angeles: The Getty Research Institute, 2008.
- Golumbia, David. "Postcolonial Studies, Digital Humanities, and the Politics of Language." Uncomputing (blog), May 31, 2013, www.uncomputing.org/?p=241.
- Gwara, Scott. *Otto Ege's Manuscripts: A Study of Ege's Manuscript Collections, Portfolios, and Retail Trade*. Medieval Manuscript Collections and Collectors in North America. Cayce: De Brailles, 2013.
- Hargittai, Eszter. "Digital Natives or Digital Naives? The Role of Skill in Internet Use." Presentation at CNI: Coalition for Networked Information, Washington, DC, December 9–10, 2013.

- Harpring, Patricia. *Introduction to Controlled Vocabularies: Terminology for Art, Architecture, and Other Cultural Worlds*. Los Angeles: J. Paul Getty Research Institute, 2010. www.getty.edu/research/publications/electronic_publications/intro_controlled_vocab/what.pdf.
- Haslhofer, Bernhard, and Wolfgang Klas. "A Survey of Techniques for Achieving Metadata Interoperability." *ACM Computing Surveys* 42 (2010): 1–37.
- Hoccleve, Thomas. *'My Complainte' and Other Poems*, edited by Roger Ellis. Exeter: University of Exeter Press, 2008.
- Hodge, Gail. *Metadata Made Simpler: A Guide for Libraries*. Bethesda: NISO Press, 2001.
- Holsinger, Bruce. "Of Pigs and Parchment: Medieval Studies and the Coming of the Animal." *PMLA* 124 (2009): 616–23.
- Horman, William. *Vulgaria uiri doctissimi Guil. Hormani Caesariburgensis*. London: Richard Pynson, 1519.
- Kiss, Farkas Gábor, Eyal Poleg, Lucie Doležalová, and Rafal Wójcik. "Old Light on New Media: Medieval Practices in the Digital Age." *Digital Philology* 2 (2013): 16–34.
- Koh, Adeline. "Niceness, Building, and Opening the Genealogy of the Digital Humanities: Beyond the Social Contract of Humanities Computing." *Differences: A Journal of Feminist Cultural Studies* 25 (2014): 93–106.
- Lawton, David. "Dullness and the Fifteenth Century." *ELH* 54 (1987): 761–99.
- Library of Congress, "DLC/Aquifer Summary of MODS Requirements and Recommendations Table." www.loc.gov/standards/mods/userguide/dlfaquifer.html.
- Lydgate, John. *The Fall of Princes*. 4 vols. Edited by Henry Bergen. Washington: Carnegie Institution of Washington, 1923.
- Morcos, Hannah. "Piecing Together the Puzzle of the Hungerford Hours." British Library Medieval Manuscripts Blog (blog), November 17, 2015. <http://blogs.bl.uk/digitisedmanuscripts/2015/11/piecing-together-the-puzzle-of-the-hungerford-hours.html>.
- National Information Standards Organization (NISO). "Understanding Metadata." Bethesda: NISO Press, 2004.
- Nikolova-Houston, Tatiana, and Ron Houston. "Building the Virtual Scriptorium." In *Digitisation Perspectives*, edited by R. Rikowski, 229–44. Rotterdam: Sense Publishers, 2011.
- Oxford English Dictionary*. "interoperability."
- Parker Library On the Web. <http://parker.stanford.edu>.
- Prescott, Andrew. "'Their Present Miserable State of Cremation': The Restoration of the Cotton Library." In *Sir Robert Cotton as Collector: Essays on an Early Stuart Courtier and His Legacy*, edited by C. J. Wright, 391–454. London: British Library Publications, 1997.
- Riley, Jenn, "Seeing Standards: A Visualization of the Metadata Universe." *Jenn Riley: Digital Librarian* (blog). <http://jennriley.com/metadatamap/>.
- Rosenheck, Donna. "OCLC: From an Historical Perspective." *The Katharine Sharpe Review* 4 (Winter 1997). <http://hdl.handle.net/2142/78252>.
- Rudy, Kathryn M. "Dirty Books: Quantifying Patterns of Use in Medieval Manuscripts Using a Densitometer." *Journal of Historians of Netherlandish Art* 2 (2010). doi:10.5092/jhna.2010.2.1.1.

- Rundle, David. "Virtual Manuscripts and the Real World. Part I." *Bonae litterae* (blog), July 8, 2013. <https://bonaelitterae.wordpress.com/2013/07/08/virtual-manuscripts-real-world-part-i/>.
- Ryder, Michael L. "Parchment: Its History, Manufacture, and Composition." *Journal of the Society of Archivists* 2 (1964): 391–99.
- Sadler, Bess, and Chris Bourg. "Feminism and the Future of Library Discovery." *Code4Lib Journal* 28 (April 2015): 1–5. <http://journal.code4lib.org/articles/10425>.
- Saenger, Paul. *A Catalogue of the Pre-1500 Western Manuscript Books at the Newberry Library*. Chicago: University of Chicago Press, 1989.
- Schmitt, Desmond. "Towards an Interoperable Scholarly Edition." *Journal of the Text Encoding Initiative* 7 (2014). <https://jtei.revues.org/979>.
- Searle, Eleanor. "Possible History." *Speculum* 61 (1986): 779–86.
- Shakespeare, William. *The Tempest*. Edited by Barbara A. Mowat and Paul Werstine. Folger Digital Texts. Washington: Folger Shakespeare Library: www.folgerdigitaltexts.org/html/Tmp.html.
- Soulah, Mohammed Ourabah, and Mohamed Hassoun. "A TEI P5 Manuscript Description Adaptation for Cataloguing Digitized Arabic Manuscripts." *Journal of the Text Encoding Initiative* 2 (2012). <https://jtei.revues.org/398>.
- Summit, Jennifer. *Memory's Library: Medieval Books in Early Modern England*. Chicago: Chicago University Press, 2008.
- TEI: Text Encoding Initiative, P5: Guidelines for Electronic Text Encoding and Interchange. Version 3.2.0. "Chapter 10: Manuscript Description." www.tei-c.org/release/doc/tei-p5-doc/en/html/MS.html.
- The Digital Walters. "Describing Manuscripts with TEI." <http://thedigitalwalters.org/Supplemental/ManuscriptDescription.html>.
- . W.169, the Hours of Pastor Denys. Walters Art Museum. www.thedigitalwalters.org/Data/WaltersManuscripts/html/W169/.
- Thesaurus for Graphic Materials*. Library of Congress. www.loc.gov/pictures/collection/tgm/.
- Treharne, Elaine. "Fleshing Out the Text: The Transcendent Manuscript in the Digital Age." *postmedieval* 4 (2013): 465–78.
- Unsworth, John. "Computational Work with Very Large Collections: Interoperability, Sustainability, and the TEI." *Journal of the Text Encoding Initiative* 1 (2011). <https://jtei.revues.org/215>.
- . "Digital Humanities Centers as Cyberinfrastructure." Presentation at Digital Humanities Centers Summit, National Endowment for the Humanities, Washington, DC, April 12, 2007.
- What the OCLC Online Union Catalog Means to Me: A Collection of Essays*. Dublin, OH: OCLC Online Computer Library Center, Inc., 1997.

DIGITAL REPRESENTATIONS OF THE PROVENANCE OF MEDIEVAL MANUSCRIPTS

TOBY BURROWS

Scope

THE FOCUS OF this chapter is on metadata for recording the provenance of medieval manuscripts, considered primarily as movable physical cultural heritage objects or artefacts. By “provenance,” I mean the stages in the history of these manuscripts from the point of their production to the present day—often spanning a period of 1,000 years or more. Among the main elements to be included in this provenance history are: who created these manuscripts, who owned them, how they were transmitted from one owner or custodian to another and from one location to another, and how and why these changes occurred.¹

In this context, provenance is an umbrella term covering a series of stages in the history of a cultural heritage object, from its creation to its current status. This history is often fragmentary, since evidence is often lacking, especially for the period before the nineteenth century. But even a fragmentary history of this kind is vital for understanding the processes by which tens of thousands of manuscripts have survived over many centuries to play a crucial role in our knowledge of medieval and early modern Europe. They serve as witnesses to a very different and distant era, which still forms a major component of the European cultural heritage.

Provenance has a different meaning in the domain of computer science: “information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness.”² Modelling provenance in this context focuses on tracking the production of digital data and digital assets, and includes such elements as file formats, software, and hardware. A useful example is given in the PROV Model Primer, which is an introduction to the PROV Data Model and Ontology.³ A blogger notices an apparent error in a chart published in an online newspaper article about crime statistics. The chart is based on data provided by a government portal, aggregated by geographical regions. The blogger needs to investigate the provenance of the various elements involved, to see whether the error lies in the chart, the article, or the government statistical data.

This is similar to the way in which the term “provenance” is used in a field like art history. Here, the emphasis is on establishing the authenticity of the object or the record,

1 Pearson, *Provenance Research*.

2 Moreau and Missier, eds., “The PROV Data Model W3C Recommendation.”

3 Gil and Miles, “The PROV Model Primer.”

through an analysis of the trail of evidence relating to its production, history, and ownership. Provenance is also a fundamental principle in archival science, where it focuses on the origin and creation of records and is used to guide the way in which records are retained and controlled: records of different provenance should be kept separate, while those with a common provenance should be kept together. There has been much debate in recent years about this concept, with ideas such as “parallel provenance” and “societal provenance” being advanced to suggest a broader, historical framework for thinking about archival collections.⁴

Provenance studies of manuscripts are also aimed, at least in part, at establishing authenticity and verifying origins. But the history of the ownership and transmission of these kinds of objects is a much broader topic, which can reveal a great deal about the nature and significance of collecting, both private and public.⁵ It can help to explain the ways in which attitudes to the past—and especially to cultural heritage—have changed over many centuries. It can also illuminate the networks of relationships between the owners of these often very valuable objects. The provenance of these objects tells us a great deal about the history, significance, and transmission of culture. A notably successful popular example of the use of provenance in this way is Neil McGregor’s *A History of the World in 100 Objects*, which reconstructs the individual histories of 100 objects now in the British Museum.⁶

The work on modelling provenance metadata presented here draws on a specific use case: a research project funded by the European Union that studies the provenance of manuscripts formerly in the collection of Sir Thomas Phillipps (1792–1872). This was the largest private collection of manuscripts and documents ever assembled, containing at least 40,000 items. Phillipps collected comprehensively rather than selectively, from a wide variety of sources and geographical areas. The collection was dispersed after his death, in a process spanning more than a century, and the manuscripts now form important and valuable components of public and private collections across the world.

Provenance is central to this project, which focuses on tracking, analyzing, and visualizing the individual histories of these manuscripts over the centuries. I want to be able to study Phillipps’s approach to collecting: what methods he used to acquire his manuscripts, where and when he bought them, and from whom, who the previous owners had been, and how far the manuscripts had travelled from their places of origin. I want to be able to study how the Phillipps manuscripts were dispersed: where and when they were sold, and to whom, how many times they have been re-sold, and where they are now. I want to carry out quantitative analyses of the scale of the collecting networks and sales activities involved in the history of the Phillipps manuscripts. I would also like to construct visualizations of manuscript histories, either on a map or as a diagram of network connections.

4 Piggott, *Archives and Societal Provenance*.

5 Dondi, “The Integration of Provenance Data,” 100–2.

6 McGregor, *A History of the World in 100 Objects*.

There is no single site or source where the materials relating to the history of the Phillipps collection are assembled. Instead, my project requires the aggregation of very varied and voluminous evidence from many different sources: inscriptions in the manuscripts themselves; sales and auction catalogues; acquisition and disposal records of individual collectors and institutions; catalogue records (printed, hand-written, or digital) associated with specific private and institutional owners; and other external documentation, including references in other documents.

Representing all this evidence in a computational setting is far from straightforward. For the Phillipps project, I reviewed existing approaches as reflected in the catalogues and databases of libraries and museums. I also examined approaches built on event-based ontologies and data models. The conceptual and practical limits of these approaches led me to a consideration of more flexible alternatives based on property graphs.

Provenance in Library Catalogues

Many surviving manuscripts are held in libraries, and libraries have had lengthy experience in managing manuscript collections. While the Machine Readable Metadata Standards (MARC) metadata format revolutionized library cataloguing in the late 1960s, it did little for manuscripts. Libraries have tended to catalogue manuscripts as if they were similar to printed books, forcing them to fit into the Procrustean bed of the MARC record structure. The unique nature of handwritten manuscripts as cultural heritage objects is a poor fit for the standardized, mass-produced descriptions of printed books for which MARC was developed.

Manuscript-oriented extensions of MARC and the Anglo-American Cataloguing Rules were subsequently developed,⁷ but these focused mainly on the intricacies of describing a medieval manuscript volume. Provenance was handled through the use of “Notes” fields. In the MARC format, there are two types of notes that are of particular relevance:

541: Immediate source of acquisition

561: Provenance note

These are unstructured fields for narrative information related to acquisition and provenance.

This narrative, note-based approach reflects that of the printed manuscript catalogues that preceded MARC and that are still being published today. These catalogues almost invariably record provenance as a series of narrative statements about ownership. A typical example looks like this:

395v. has “394” in a 15c. hand and in a later hand are three Benedictions; Br. by a 19c. hand is “Missale Romanum. Codex missalis A.B.” and autograph “Payne” (Messrs Payne and Foss from whom Sir Thomas Phillipps purchased the volume in 1848); inside front cover is a green seal with lion rampant (device

⁷ Pass, *Descriptive Cataloging*.

```

<history>
  <origin notBefore="1225" notAfter="1275">Written in Spain or Portugal in the middle of the 13th
    century (the date 1042, given in a marginal note on f. 97v, cannot be correct.)</origin>
  <provenance>The Spanish scholar <name type="person">Benito Arias Montano</name> (1527-1598) has
    written his name on f. 97r, and may be presumed to have owned the manuscript. </provenance>
  <provenance>It came somehow into the possession of <foreign xml:lang="da">etatsråd</foreign>
    <name type="person">Holger Parsberg</name> (1636-1692), who has written his name twice, once
    on the front pastedown and once on f. 1r, the former dated <date>1680</date> and the latter
    <date>1682</date>. </provenance>
  <provenance>Following Parsberg's death the manuscript was bought by <foreign>etatsråd</foreign>
    <name type="person">Jens Rosenkrantz</name> (1640-1695) when Parsberg's library was
    auctioned off (23 October 1693). </provenance>
  <acquisition notBefore="1696" notAfter="1697">The manuscript was acquired by Árni Magnússon from
    the estate of Jens Rosenkrantz, presumably at auction (the auction lot number 468 is written
    in red chalk on the flyleaf), either in 1696 or 97. </acquisition>
</history>

```

Figure 9.1. Example of multiple <provenance> elements in the TEI Guidelines.

of Philipps) and "12289 Ph." (no. repeated on spine); the MS. was acquired in 1910 by A. G. Little, then later by W. H. Robinson, and was subsequently presented to the library (details of the 1949 negotiations that resulted in the donation can be found in the article by Fr C. Kelly OFM, "Franciscan Scholarship in the Middle Ages—II", *Catholic Review*, Vol.V (1949), pp. 213–14)

This account of a thirteenth-century Franciscan Missal owned by St. Paschal's College in Melbourne (currently on loan to the State Library of Victoria) records the evidence relating to provenance in chronological order, assembled into a single narrative paragraph.⁸

This narrative approach to recording manuscript provenance is reflected in the Text Encoding Initiative (TEI). The TEI's MS-Desc (Manuscript Description) section contains a <history> element for encoding information from a catalogue entry about the lifecycle of a manuscript.⁹ There are three components, defined as follows:

1. <origin>: descriptive or other information about the origin of a manuscript or manuscript part;
2. <provenance>: descriptive or other information concerning a single identifiable episode during the history of a manuscript or manuscript part, after its creation but before its acquisition;
3. <acquisition>: descriptive or other information concerning the process by which a manuscript or manuscript part entered the holding institution.

Multiple <provenance> elements can be used to record distinct periods of ownership, as the example in Figure 9.1 from the TEI Guidelines demonstrates:

As the example shows, this approach is designed primarily for identifying specific semantic elements within the narrative histories given in printed manuscript catalogues.

⁸ Sinclair, *Descriptive Catalogue*, no. 181.

⁹ Text Encoding Initiative, P5, "Guidelines."

Library catalogues sometimes use provenance notes to derive additional access points for persons and organizations. The provenance-oriented database “Material Evidence in Incunabula (MEI),” produced by the Consortium of European Research Libraries (CERL), implements this approach in a systematic way.¹⁰ MEI aims to incorporate copy-specific information about incunabula (pre-1500 printed books), which provides evidence relevant to the provenance and use of that copy. This includes ownership notes, which may be categorized as private or institutional, religious or lay, female or male, and by profession. Persons and places are linked to the CERL Thesaurus in order to enable biographical and geographical analysis. Over 20,000 people in the CERL Thesaurus are linked to catalogue records for the books that they once owned.

The additional metadata required for these kinds of provenance information are added to the standard MARC record structure, either as notes or as corporate and personal name access points. Searches in MEI can be qualified by using provenance elements, including time and place, former owner (gender, status, profession, role), method of acquisition, and nature of evidence. Provenance-related searches in the catalogues of the contributing libraries are dependent on specific functionality, but are likely to be limited to person or organization.

A similar approach has been implemented in the catalogues of a number of major research libraries. In Harvard University’s HOLLIS catalogue, for example, a personal name access point for former owners is usually added to the catalogue record for a specific manuscript. This means that a search for “Phillipps, Thomas, Sir, 1792–1872, former owner” returns 355 records for manuscripts once owned by Phillipps and now owned by Harvard.

More recent developments in library-related metadata schemas do little to address the representation of provenance. The Functional Requirements for Bibliographic Records (FRBR) offers a metadata model designed to supersede MARC, based on relationships between entities (persons, places, and things) rather than on fields in a catalogue record. But there does not appear to be any specific modelling of provenance in FRBR. Although one of the core entities is “Event”, defined as “an action or occurrence”, the scope note for this entity makes it clear that FRBR is only interested in events insofar as they are the *subject* of a bibliographic work.¹¹

FRBR formed one of the main building blocks for the development of the new RDA (Resource Description and Access) metadata standard for the library sector. RDA contains elements designed to specify the owner, custodian, and “other person, family, or corporate body associated with an item.”¹² It also provides a series of properties that can be used to specify the relationship between these people and a specific item:

- Has Custodial History
- Has collector

10 Dondi, “Provenance Records,” 15–19.

11 *Functional Requirements for Bibliographic Records*.

12 Resource Description and Access, Element Analysis Table.

- Has depositor
- Has donor
- Has seller
- Has annotator
- Has inscriber
- Has current owner
- Has former owner
- Has custodian
- Has owner
- Has immediate source of acquisition.

But this model is still centred on the object, and only serves to link people or organizations with specific items. The “Custodial History” is reminiscent of a note in the catalogue record, and there is no modelling of provenance as such.

Provenance is also important for museums and galleries. Their databases have tended to follow the same general approach as library catalogues. Provenance information is generally included in the record for an individual object, in the form of a chronologically organized note. Museum databases, like their library equivalents, may use provenance notes to derive additional access points for persons and organizations as owners or custodians of an object. The American Association of Museums provides a standard for structuring provenance notes in the *AAM Guide to Provenance Research*.¹³

A good example of this approach in practice can be found in the collection database of the National Gallery of America. Users are able to search by “provenance name,” leading to a biographical record for each owner and a list of objects associated with that owner. The “provenance” section of an object record follows the structure recommended in the *AAM Guide*:¹⁴

The provenance for a work of art in the National Gallery’s collection is listed in chronological order, beginning with the earliest known owner. Life dates, if known, are enclosed in brackets. The names of dealers, auction houses, or agents are enclosed in parentheses to distinguish them from private owners. Relationships between owners and methods of transactions are indicated by punctuation: a semicolon is used to indicate that the work passed directly between two owners, and a period is used to separate two owners if a direct transfer did not occur or is not known to have occurred. Footnotes are used to document or clarify information.

For each object, this approach produces a provenance field consisting of a chronological listing of owners, often with footnotes citing relevant evidence. A typical record looks like this (for *The Annunciation*, by Jan van Eyck):

13 Yeide, Akisha, and Wash, *AAM Guide to Provenance*.

14 National Gallery of Art, *How to Read Gallery Provenance Texts*.

Possibly the Chartreuse de Champmol, near Dijon.[1] Sale, Paris, 1819. (Charles J. Nieuwenhuys, Brussels). William II, King of the Netherlands [d. 1849], in Brussels until 1841, thereafter The Hague;[2] (sale, The Hague, 12 August 1850, no. 1); bought by Bruni for Czar Nicholas I of Russia [d. 1855], Saint Petersburg; Imperial Hermitage Museum, Saint Petersburg;[3] purchased June 1930 through (Matthiesen Gallery, Berlin; P. & D. Colnaghi & Co., London; and M. Knoedler & Co., New York) by Andrew W. Mellon, Pittsburgh and Washington, D.C.; deeded 5 June 1931 to The A.W. Mellon Educational and Charitable Trust, Pittsburgh; gift 1937 to NGA.

The Carnegie Museum of Art has developed a draft *Provenance Standard* that aims to provide a more rigorous rendering of the information in the *AAM Guide*, making it more suitable for machine-based parsing.¹⁵ This is incorporated into the Museum's Elysa software for entering provenance lists of this kind. From this structure, geographical and chronological visualizations of an object's history can be generated.¹⁶ The underlying data model is chronological. It begins with a Timeline for each object, which contains a number of Periods. Each Period includes a Party, an Acquisition Method, a Location, and a Beginning and End.

A mapping document or crosswalk produced by the Getty Research Institute compares the treatment given to provenance by various metadata schemas from the museum sector. The CDWA (Categories for the Description of Works of Art) Lite schema, widely used until it was superseded by LIDO (see below), is fairly typical. It contains only generic unstructured elements for "Acquisition Source" and "Custodial History." There is a separate section of the Getty document for "Creating and commissioning," but this also contains relatively little detail or internal structure.¹⁷

In general, though, library catalogues and museum databases are all about discoverability: finding specific manuscripts and objects held in specific institutional collections. The focus is on providing the current location and identifying the main characteristics of the item involved. From this point of view, the history of the object can be briefly recounted in a note—if it is recounted at all. There is no support for analytical questions of the kind that are central to projects like mine.

A different approach is provided by the Schoenberg Database of Manuscripts, managed by the University of Pennsylvania Library.¹⁸ This database collects evidence relating to the movements of medieval and Renaissance manuscripts, drawn from sales and auction catalogues, and collection catalogues. Each piece of evidence is given a separate record, which includes a "provenance information" list of former owners and dates (where this information is available in the source). Entries thought to relate to the same manuscript are linked to a "master record" for that manuscript, and this displays all the various provenance information fields from the linked entries. A combined provenance

¹⁵ Carnegie Museum of Art, CMOA Provenance Standard.

¹⁶ Berg-Fulton, Newbury and Snyder, "Art Tracks."

¹⁷ Baca and Harpring, *Categories for the Description of Works of Art*.

¹⁸ Schoenberg Database of Manuscripts.

list can then be constructed manually, and contributed as a “personal observation” linked to the master record.

The Schoenberg Database does enable some aggregated analysis of all this provenance information. Browsing by the “provenance” facet enables you to see all the major provenance actors since the seventeenth century, headed by Sir Thomas Phillipps with nearly 20,000 entries. It is possible to combine this with other criteria—such as place and date—to come up with a list of sixty-seven entries relating to Italian fourteenth-century manuscripts that were once owned by both Phillipps and the Italian collector Luigi Celotti. The Schoenberg Database is an important and valuable source of data for my Phillipps project, but it has some key limitations: it does not cover post-Renaissance manuscripts, it does not include manuscripts with fewer than five leaves, and it does not enable visualizations. Only in a limited number of cases does it combine the provenance information relating to a single manuscript, since this has to be done manually.

In the case of the Phillipps manuscripts, there are also a number of specialized handwritten provenance records. The British Library has a set of thousands of index cards, each of which contains handwritten provenance statements relating to a single Phillipps manuscript. Information was added to these cards over several decades up to the 1990s, but they have never been digitized. Another specific resource, held in the Bodleian Library, consists of handwritten annotations interleaved into a photocopy of the printed catalogue of Phillipps manuscripts. It provides provenance statements for transactions relating to these manuscripts during the twentieth century. The Cambridge University Library also has an annotated catalogue of this kind.

Event-Based Modelling of Provenance

Both library catalogues and museum databases consist of records for individual objects, which usually have their provenance recorded in the narrative histories and notes attached to each record. But, as these notes make clear, provenance actually consists of a series of events (such as sales) and situations (such as periods of ownership). Representing and visualizing provenance in its fullest sense requires a data model built around events and situations rather than descriptions of objects.

The concept of an event plays an important role in several data models and ontologies that are used for the description of cultural heritage objects. Comité international pour la documentation-Conceptual Reference Model (CIDOC-CRM) is the most ambitious of these. Developed by the museum community as “a formal ontology intended to facilitate the integration, mediation, and interchange of heterogeneous cultural heritage information,” it is now an ISO standard (ISO21127:2006). The CIDOC-CRM ontology is relatively large, with 86 classes and 137 unique properties.¹⁹

CIDOC-CRM models events as temporal entities. “Event” (E5) is a subclass of “Period” (E4), which is itself a subclass of “Temporal Entity” (E2). “Period” is defined as “a set of coherent phenomena or cultural manifestations bounded in time and space.” “Event” is

19 Crofts, Doerr, Gill, Stead, and Stiff, eds., “Definition of the CIDOC.”

defined as “changes of states in cultural, social or physical systems, regardless of scale, brought about by a series or group of coherent physical, cultural, technological or legal phenomena.” The accompanying scope note observes: “the distinction between an E5 Event and an E4 Period is partly a question of the scale of observation. Viewed at a coarse level of detail, an E5 Event is an ‘instantaneous’ change of state. At a fine level, the E5 Event can be analyzed into its component phenomena within a space and time frame, and as such can be seen as an E4 Period.” The E5 entity class sits at the top of a hierarchy of more specific types of events (E6, E7, E67, and so on), which are illustrated by the following examples:

- The destruction of Herculaneum by volcanic eruption in 79 AD (E6—Destruction)
- World War II (E7—Activity)
- The birth of Cleopatra (E67—Birth).

There is a recent “provenance” extension of CIDOC-CRM, known as CRM_{dig}²⁰ but this relates to the provenance of digital objects, not to the history of cultural heritage materials. Provenance in the latter sense can be represented in CIDOC-CRM by combining “Event” entities and properties to form an event-centred framework. Christian-Emil Ore has discussed how this might work in practice, using a medieval manuscript—the Kringla Leaf—as his example.²¹ This leaf has a fairly complicated history, summarized in twelve steps by Ore. The CIDOC-CRM representation of this history is even more complex, involving thirty-eight entities and forty-three different relationships between them. Fifteen of the entities are provenance events, and eight of them are “information carriers” (i.e., versions of the manuscript). The three diagrams summarizing this representation are dense and overlapping, and hard to read and grasp. This suggests that, although CIDOC-CRM can indeed represent the provenance of a manuscript, the complexity of the results makes it hard to envisage their implementation in a practical setting.

The CIDOC-CRM approach to events is also found in Lightweight Information Describing Objects (LIDO), the lightweight museum schema that has replaced Categories for the Description of Works of Art (CDWA) Lite.²² The “Event” element in LIDO comes with an extensive list of types: Acquisition, Creation, Finding, Modification, Use, Collecting, Designing, Destruction, Excavation, Exhibition, Loss, Move, Order, Part addition, Part removal, Performance, Planning, Production, Provenance, Publication, Restoration, and Transformation. The meaning of “Provenance” in this context is unclear; is it just a general term for events which do not fit into one of the more specific categories? The idea that provenance is a type of event—rather than the combination of a series of events—seems to be conceptually confused.

CIDOC-CRM has been influential beyond the museum community, with a number of other sectors working to link data models and ontologies to it. For libraries, this is the main purpose behind the development of FRBR_{oo}—a version of FRBR expressed in terms

²⁰ Theodoridou et al., “Modeling and Querying,” 169–210.

²¹ Ore, “CIDOC CRM and Object Provenance.”

²² Coburn et al., “LIDO—Lightweight Information Describing Objects.”

of entities and their relationships. FRBR_{oo} contains an entity called F8—Event, based on CIDOC-CRM’s E4—Period.²³ FRBR_{oo} quotes the CIDOC-CRM scope note, and gives the following examples:

- The battle of Trafalgar
- Printing for the publisher named Doubleday in 2003 all the copies of the first print run of the novel titled *Da Vinci Code*
- Having the initial idea that eventually resulted in the existence of the opera titled *Der fliegende Holländer*
- Creating for Mozart’s 41st Symphony the uniform title that was thereafter consistently used to refer unambiguously to that symphony everywhere in the Library of Congress’s catalogue.

These examples are consistent with the “changes in state” approach found in CIDOC-CRM. FRBR_{oo} also adds the comment that “this class contains sets of coherent phenomena or cultural manifestations bounded in time and space,” and gives Impressionism as an example. This introduces a certain amount of confusion into the concept of “Event” by conflating CIDOC-CRM’s definitions of “Period” and “Event.” The example is also confusing. Impressionism is not simply a change of state or even a temporal period; in fact, it is hard to see how Impressionism could be called an event in the CIDOC-CRM sense.

The Europeana Data Model (EDM) used by the Europeana Digital Library also contains a core entity class for “Event.” As the definition in Table 9.1 makes clear, EDM follows FRBR_{oo} in combining the definitions of “Event” and “Period” from CIDOC-CRM.²⁴

The EDM is intended to enable event-centred networks to be attached to any given object. Events can be treated as “hubs” that relate an object to other, directly connected entities. These relationships are represented using three basic properties:

- edm:wasPresentAt, holding between any resource and an event it is involved in
- edm:happenedAt, holding between an event and a place
- edm:occurredAt, holding between events and the time spans during which they occurred.

Different types of events are not specified in the EDM Data Model; instead, they are identified by the data provider or when Europeana enriches a record with other vocabularies. This results in a simplified version of the provenance modelling undertaken by Ore with CIDOC-CRM. The *EDM Data Model Primer* provides the example shown in Figure 9.2.²⁵

In addition to following CIDOC-CRM and FRBR_{oo}, Europeana’s approach to events draws on ABC Harmony, an earlier attempt to build an ontology for integrating

²³ Bekiari, Doerr, and Le Bouef, eds., “FRBR: Object-oriented Definition,” 35–36.

²⁴ Europeana, “Definition of the Europeana Data Model,” 12–13.

²⁵ Isaac, ed., *Europeana Data Model Primer*, 15–17.

Table 9.1. The definition of the “Event” class in the Europeana Data Model.

Class name: edm:Event	
URI	http://www.europeana.eu/schemas/edm/Event
Label	Event
Definition	An event is a change “of states in cultural, social, or physical systems, regardless of scale, brought about by a series or group of coherent physical, cultural, technological, or legal phenomena” (E5 Event in CIDOC-CRM) or a “set of coherent phenomena or cultural manifestations bounded in time and space” (E4 Period in CIDOC-CRM)
Subclass of	edm:NonInformationResource
Equivalent class	E4_Period (CIDOC-CRM), Event (FRBR), Temporality (ABC Harmony)
Obligation & Occurrence	
Example	• The act of painting Mona Lisa • The 2nd World War • The change of custody of Mona Lisa
Comment	Events are identified either by the content provider or by Europeana enrichment at ingestion time. This class is a domain of edm:happenedAt and the domain of edm:occurredAt

metadata in digital libraries. ABC Harmony specifically included provenance among its use cases:²⁶

The ABC model allows users to ask much more sophisticated queries than is possible via less expressive metadata models such as Dublin Core e.g., “Tell me all of the previous owners of an object,” “Give me all of those objects which were acquired as gifts and the donor’s name and address.” By using the ABC model, one is able to record and retrieve the history of an object from its creation, through to its use, change of ownership, relocation, modification, digitization and repurposing.

Like CIDOC-CRM, ABC Harmony modelled Event as a subclass of Temporality, but it also included Action as another subclass and added a subclass for Situation, which was intended to cover time periods during which the properties of an object were static. Some subsequent work was done on mapping between ABC Harmony and CIDOC-CRM.²⁷

²⁶ Lagoze and Hunter, “The ABC Ontology and Model,” 8–9.

²⁷ Doerr, Hunter, and Lagoze, “Towards a Core Ontology.”



Figure 9.2. Amphora of Tuthmosis III—an event-centric description (without aggregation and related digital representation entities).

This produced a class for Event with a subclass for Action (ABC) or Activity (CIDOC-CRM). But it proved more difficult to reconcile ABC's use of the Situation subclass, given that CIDOC-CRM did not recognize this type of temporality.

There have been several other recent initiatives designed to develop models and ontologies for events. They are all intended to be relatively simple and generic, with a small number of classes and a limited number of properties. They include LODE: an Ontology for Linking Open Descriptions of Events (one class, six properties)²⁸ and The Event Ontology (four classes, seventeen properties).²⁹ Also worth noting is the Simple Event Model (SEM),³⁰ which is being used in the context of research aimed at the automatic extraction of historical event information from unstructured texts.

The ontologies and data models examined here generally define a class of entities called "event" and attribute various properties to it. For them, events are nameable time-based actions or activities, which occur within a broader period of time. In most cases, they are insufficiently detailed to serve as a basis for structuring provenance information. In the case of CIDOC-CRM, though, the approach to modelling provenance histories of manuscripts may become too complex to be practical and workable.

It is worth noting that the nature of "events" is a recognized contemporary philosophical problem. An important contributor to this field is Peter Gärdenfors, who puts the problem of events at the centre of his recent work on semantics and conceptual spaces. He argues that this kind of ontological approach is far too narrow for modelling events in a computational setting.³¹ Events, in his view, should not be treated as nameable entities with specific properties, which can be manipulated by symbolic logic and deductive reasoning. Instead, he envisages conceptual spaces where events are formed from the intersection of different agents, patients, forces, and properties. His work is still largely theoretical, though, and does not yet offer a basis for a different type of computational representation.

Beyond Conceptual Modelling

While there is no current representational framework for the "conceptual spaces" envisaged by Gärdenfors, it may still be possible to find a way of modelling events that goes beyond the simplicity of treating all events as nameable entities. One step in this direction might be the property graph.³² The structure of nodes and relationships (or vertices and edges) found in property graphs can link together the various actors in an event without having to create an "event" entity as such. The nature of the connections between actors in an event can be expressed as relationships between nodes. A relationship must

28 LODE.

29 The Event Ontology.

30 Van Hage et al., "Design and Use of the Simple Event Model (SEM)," 128–36.

31 Gärdenfors, *The Geometry of Meaning*.

32 Robinson, Webber, and Eilfrem, *Graph Databases*.

always have a start node and an end node, and must have a direction. Both nodes and relationships can contain properties.

This approach can be applied to modelling the provenance of manuscripts. A typical example of a provenance statement looks like this:

In 1862, Sir Thomas Phillipps bought Phillipps MS 16402 in London as part of the Sotheby's sale of the collection of Guglielmo Libri.

This statement can be analyzed into the components of a property graph as shown in Table 9.2. There is also an implied relationship between Libri and the manuscript, as owner and consignor.

A relatively simple initial data model can be derived from this kind of analysis, for implementation in a property graph database using software like Neo4j.³³ Four basic nodes—Agent, Object, Publication and Work—can be identified. “Object” is primarily a physical entity: the manuscript volume itself. “Work” is a conceptual entity—the text carried by a manuscript, or the description of a manuscript contained in a catalogue (like FRBR’s “Work”). “Agent” can be given the types “Person” and “Organization.” Relationships can include the basic set of transactions relating to a manuscript: “Bought,” “Sold,” “Produced,” and so on. These can be reciprocal relationships: a Person bought a Manuscript, and the Manuscript was bought by the Person. Table 9.3 shows some of the relationships that are possible between the various types of nodes.

In this kind of model, a provenance event like the Phillipps example given above is represented by a group of nodes and their relationships, together with the properties specifying the time and place at which these relationships occurred. This approach is hospitable to periods as well as to specific times: a Person owned a Manuscript, with a property attached to the “owned” relationship for a time period rather than a specific date. This can also apply to uncertain dates: an Organization produced a Manuscript, and the “produced” relationship has an attached property for possible start and end dates.

A similarly flexible approach to modelling provenance is offered by software called nodegoat.³⁴ Developed in the Netherlands by the Lab1100 group for use in research projects mapping correspondence networks in early modern Europe, nodegoat uses some of the same ideas as property graphs, without being a property graph database. The basic nodegoat structure consists of objects, with sub-objects attached to them. These sub-objects can embed relationships to other objects, as well as containing properties like time and place.

Applying this framework to manuscript provenance involves using sub-objects to model event relationships as shown in Table 9.4. A “Manuscript” object can have sub-objects for transactions such as “Sold,” “Donated,” “Owned,” and so on. These sub-objects can contain links to the Persons and Organizations participating in each transaction, as well as the times and places where the transaction occurred. These times can be periods as well as specific dates.

³³ Van Bruggen, *Learning Neo4j*.

³⁴ Van Bree and Kessels, “Mapping Memory Landscapes in nodegoat,” 274–78.

Table 9.2. The components of a property graph for the purchase of Phillipps MS 16402.

1862	Property of “Bought” relationship
Sir Thomas Phillipps	Node
Bought	Relationship
Phillipps MS 16402	Node
London	Property of “Bought” relationship
Sotheby’s	Node
Sale	Relationship
Guglielmo Libri	Node

Table 9.3. Sample nodes and relationships for Neo4j property graph.

Node	Relationship	Node
AGENT: Person	BOUGHT	OBJECT: Manuscript
AGENT: Organization	SOLD	OBJECT: Manuscript
OBJECT: Manuscript	CONTAINS	WORK: Text
AGENT: Person	COMPOSED	WORK: Text
PUBLICATION: Catalogue	CONTAINS	WORK: Description
OBJECT: Manuscript	DESCRIBED_AS	WORK: Description
AGENT: Organization	PRODUCED	PUBLICATION: Catalogue
WORK: Exhibition	DESCRIBED_IN	PUBLICATION: Catalogue

Table 9.4. Sub-objects of the “Manuscript” Object in the nodegoat framework.

OBJECT	SUB-OBJECT	RELATED TO
Manuscript	Sold Donated Owned Described In Produced Contains	Person, Organization Catalogue Text

Table 9.5. A partial history of Phillipps MS 12264 modelled in nodegoat.

Sub-object	Date Start	Date End	Location	Other actors
Produced	775	800	Nonantola (IT)	Abbey of Nonantola
Owned	1660	1818	Roma (IT)	Basilica di Santa Croce in Gerusalemme
Sold	1848		Roma (IT)	Thomas Thorpe; Giambattista Petrucchi; Sir Thomas Phillipps
Owned	1848	1872	Cheltenham (GB)	Sir Thomas Phillipps
Sold	09-06-1924		Cheltenham (GB)	Thomas Fitzroy Fenwick; Alfred Chester Beatty
Owned	09-06-1924	1954	London (GB)	Alfred Chester Beatty
Sold	24-06-1969		London (GB)	Sotheby's; Alfred Chester Beatty; Biblioteca Nazionale Centrale di Roma
Owned	24-06-1969	open	Roma (IT)	Biblioteca Nazionale Centrale di Roma

The result is a record for each manuscript that summarizes all the events in its provenance history. Table 9.5 is a partial history of Phillipps MS 12264, as modelled in nodegoat.

Using software like nodegoat and Neo4j is a good way of implementing flexible and customized data models, which can be elaborated and developed over time. It can also produce visualizations of networks of objects and the relationships between them, including their locations in space and time. There is good support for path analysis and pattern analysis, enabling the exploration of specific but semantically complex research questions. In the case of this example, these might include:

- Show all the occasions when Phillipps bought manuscripts previously owned by Beatty
- Show all the events in which Phillipps MS12264 has been involved during the course of its history
- Show all manuscripts written in a Romance language in the fourteenth or fifteenth centuries which were acquired by Phillipps from Italian collectors

- Show the dispersal of a specific group of manuscripts from the Philipps collection, with their geographical movements over time.

Data can be exported for reuse from nodegoat and Neo4j in a variety of formats, including JSON and CSV/Excel. Developing a customized data model might be seen as a potential barrier to sharing data with services that rely on event-based ontologies and data models, including areas of the Linked Data Cloud. But this can be minimized by careful design that allows for mapping to frameworks like CIDOC-CRM or LIDO. Nodes and their types in Neo4j can be mapped to entity classes in CIDOC-CRM, and relationships can be mapped to properties. In nodegoat, object types can be mapped to entity classes, and sub-objects can be mapped to properties.

Conclusion

Property graph databases like Neo4j and software such as nodegoat avoid many of the difficulties arising from the way in which most formal ontologies and data models treat events. They offer a way of handling the conceptually fuzzy world of events that is closer to the way in which we think about them. Modelling them as the intersection of various actors, relationships, and properties is a much more fruitful and flexible approach.

The suggestions provided here are far from the final word on the problem of modelling provenance metadata for cultural heritage objects. But they do have the merit of treating provenance as a chronological series of events, while treating each event in that series as the intersection of a range of people, organizations, objects, times, and places. For projects like mine, this kind of data model offers a workable way of organizing the evidence relating to the provenance of many thousands of manuscripts. It enables me to analyze and visualize their individual and collective histories and to reconstruct, in a digital environment, what Alan Liu might call the “network archaeology” of the Philipps collection.³⁵

Bibliography

- Baca, Murtha, and Patricia Harpring. *Categories for the Description of Works of Arts*. Rev. ed. Los Angeles: Getty Research Institute, 2011. www.getty.edu/research/publications/electronic_publications/cdwa/index.html.
- Bekiar, Chrysoula, Martin Doerr, and Patrick Le Bouef, eds. “FRBR Object-oriented Definition and Mapping to FRBR,” 2008. http://archive.ifa.org/VII/s13/wgfrbr/FRBRoo_V9.1_PR.pdf.
- Berg-Fulton, Tracey, David Newbury, and Travis Snyder. “Art Tracks: Visualizing the Stories and Lifespan of an Artwork,” MW2015: Museums and the Web. <http://mw2015.museumsandtheweb.com/paper/art-tracks-visualizing-the-stories-and-lifespan-of-an-artwork/>.

35 Liu, “Remembering Networks”; Burrows, “The History and Provenance of Manuscripts.”

- Burrows, Toby. "The History and Provenance of Manuscripts in the Collection of Sir Thomas Phillipps: New Approaches to Digital Representation." *Speculum* 92 (2017): S39–S64.
- Carnegie Museum of Art. "CMAA Provenance Standard, draft version" (2015). www.museumprovenance.org/standard/.
- Coburn, Erin, Richard Light, Gordon McKenna, Regine Stein, and Axel Vitzthum, "LIDO—Lightweight Information Describing Objects Version 1.0" (2010). <http://www.lido-schema.org/schema/v1.0/lido-v1.0-specification.pdf>.
- Crofts, Nick, Martin Doerr, Tony Gill, Stephen Stead, and Matthew Stiff, eds., "Definition of the CIDOC Conceptual Reference Model, version 5.0.4" (2011). www.cidoc-crm.org/docs/cidoc_crm_version_5.0.4.pdf.
- Doerr, Martin, Jane Hunter, and Carl Lagoze. "Towards a Core Ontology for Information Integration." *Journal of Digital Information* 4 (2003). <https://journals.tdl.org/jodi/index.php/jodi/article/view/92/91>.
- Dondi, Cristina. "The Integration of Provenance Data for the Reconstruction of the Dispersed European Book Heritage." In *Safeguard of Cultural Heritage: A Challenge from the Past for the Europe of Tomorrow: COST Strategic Workshop July 11th–13th, Florence, Italy*, edited by Marco Foravanti and Saverio Mecca, 100–2. Firenze: Firenze University Press, 2011.
- . "Provenance Records in the CERL Thesaurus and in Material Evidence in Incunabula." *Sborník Národního muzea / Acta Musei Nationalis Pragae, series C—Historia Litterarum* 58 (2013): 15–19.
- Europeana. "Definition of the Europeana Data Model v5.2.7" (2014): 12–13. http://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation//EDM_Definition_v5.2.7_042016.pdf.
- Functional Requirements for Bibliographic Records: Final Report* / IFLA Study Group on the Functional Requirements for Bibliographic Records. Munich: Saur, 1998.
- Gärdenfors, Peter. *The Geometry of Meaning: Semantics Based on Conceptual Spaces*. Cambridge, MA: MIT Press, 2014.
- Gil, Yolanda, and Simon Miles. "The PROV Model Primer: W3C Working Group Note 30 April 2013." www.w3.org/TR/2013/NOTE-prov-primer-20130430/.
- Isaac, Antoine, ed. "Europeana Data Model Primer" (2013). <http://pro.europeana.eu/documents/900548/770bdb58-c60e-4beb-a687-874639312ba5>.
- Lagoze, Carl, and Jane Hunter. "The ABC Ontology and Model." *Journal of Digital Information* 2 (2001): 8–9. <https://journals.tdl.org/jodi/index.php/jodi/article/viewFile/44/47>
- Liu, Alan. "Remembering Networks: Agrippa, RoSE, and Network Archaeology." Paper presented at Network Archaeology conference, Miami University, Ohio, April 21, 2012.
- LODE. <http://linkedevents.org/ontology/>.
- McGregor, Neil. *A History of the World in 100 Objects*. London: Allen Lane, 2011.
- Moreau, Luc, and Paolo Missier, eds. "PROV-DM: The PROV Data Model W3C Recommendation 30 April 2013." www.w3.org/TR/prov-dm/

- National Gallery of Art. "How to Read Gallery Provenance Texts" (2015). www.nga.gov/content/ngaweb/Collection/collection-information.html.
- Ore, Christian-Emil Smith. "CIDOC CRM and Object Provenance" (2009). www.tei-c.org/SIG/Ontologies/presentations/CIDOC_CRM_provenance.pdf.
- Pass, Gregory. *Descriptive Cataloging of Ancient, Medieval, Renaissance, and Early Modern Manuscripts*. Chicago: Association of College and Research Libraries, 2003.
- Pearson, David. *Provenance Research in Book History*. London: British Library, 1994.
- Piggott, Michael. *Archives and Societal Provenance: Australian Essays*. London: Chandos, 2012.
- Resource Description and Access (RDA), Element Analysis Table. www.rda-jsc.org/docs/6JSC-RDA-element-analysis-table-rev.pdf.
- Robinson, Ian, Jim Webber, and Emil Eilfrem. *Graph Databases*. Sebastopol: O'Reilly, 2013.
- Schoenberg Database of Manuscripts. <http://dla.library.upenn.edu/dla/schoenberg/index.html>.
- Sinclair, Keith Val. *Descriptive Catalogue of Medieval and Renaissance Western Manuscripts in Australia*. Sydney: University of Sydney Press, 1969.
- Text Encoding Initiative, P5. "Guidelines for Electronic Text Encoding and Interchange version 2.9.1: 10.8 MS-Desc: History." October 15, 2015. www.tei-c.org/release/doc/tei-p5-doc/en/html/MS.html#mshy.
- The Event Ontology. <http://motools.sourceforge.net/event/event.html>.
- Theodoridou, Maria, Yannis Tzitzikas, Martin Doerr, Yannis Marketakis, and Valantis Melessanakis. "Modeling and Querying Provenance by Extending CIDOC CRM." *Distributed and Parallel Databases* 27 (2010): 169–210.
- Van Bree, Pim, and Geert Kessels. "Mapping Memory Landscapes in Nodegoat." In *Social Informatics*, edited by L. M. Aiello and D. McFarland, 274–78. Berlin: Springer International, 2015.
- Van Bruggen, Rick. *Learning Neo4j*. Birmingham: Packt, 2014.
- Van Hage, Willem Robert, Véronique Malaisé, Roxane Segers, Laura Hollink, and Guus Schreiber. "Design and Use of the Simple Event Model (SEM)." *Web Semantics: Science, Services and Agents on the World Wide Web* 9 (2011): 128–36.
- Yeide, Nancy H., Konstantin Akisha, and Amy L. Wash. *AAM Guide to Provenance Research*. Washington, DC: American Alliance of Museums Press, 2011.

BRIDGING THE GAP: MANAGING A DIGITAL MIEVEAL INITIATIVE ACROSS DISCIPLINES AND INSTITUTIONS

JOSEPH KOIVISTO, LILLA KOPÁR, AND NANCY L. WICKER

THE INCREASING UBIQUITY of digitally enhanced medieval research initiatives serves as testament to the “brave new world” of humanities research in which technological platforms and methodologies are closely integrated with traditional scholarship.¹ Such digital humanities approaches are also proving to be quite fruitful to researchers who seek to promote inter-institutional collaboration in ways that defy conventional classifications of disciplinary domains. Scholarship in the field of medieval studies is no longer limited to the isolated work of lone scholars within narrowly defined domains of practice. Instead, many modern medievalists now find themselves engaged in collaboration with information technology specialists, library and information science practitioners, programmers, and other digital humanities scholars distributed across departments, institutions, and time zones. As a result of this new approach to the “digital medieval project,” institutional and external granting bodies—both national and international—have expanded their funding opportunities for faculty members, researchers, and graduate students to pursue innovative research projects and develop digital tools and platforms. These funding bodies require that scholars develop data management and project management plans, yet little emphasis has been placed on the managerial function of the primary investigators in team-oriented research settings.

We (Koivisto, Kopár, and Wicker) have learned about project management as we developed Project Andvari, an international collaborative project designed to create a free digital portal that will provide integrated access to collections of northern European art and artefacts of the early medieval period (fourth to twelfth centuries). Through our work on this project, we have discovered that the inter-institutional and interdisciplinary nature of effective digital medieval projects presents a bevy of project management complications that represent considerable roadblocks to the planning, development, and implementation of proposed research initiatives. Proper project management approaches, custom-tailored to the needs of interdisciplinary medieval initiatives, must be implemented in order to avoid unnecessary complications or project failure. But what resources are at our disposal? Despite the acknowledged importance of project management, domain-specific approaches and resources do not occupy a primary position in overall research considerations.² In a scholarly environment in which collaborative

¹ Svensson, “The Digital Humanities as a Humanities Project.”

² For instance, the “Project Management” page of the Digital Medievalist Wiki was once a sparsely documented resource for medievalists seeking guidance on managing digital humanities initiatives, but it no longer exists outside of the Internet Archive.

initiatives hold great significance for research, scholarship, and career outlooks,³ we can hardly afford to leave project management as an unacknowledged dimension of digital medieval initiatives.

From initial proposal writing to the management of grant funds, reporting, and ongoing project development, project management—in the form of direct resource and personnel management as well as the management of institutional expectations—has a tangible influence on day-to-day project activities. Concerns over appropriate allocation of resources, governance of external grant funds and institutional investment in cost-sharing, and distribution of appropriate organizational credit for successful (or failed) initiatives arise at every project stage and may potentially derail otherwise well-founded projects. In addition, inter-institutional initiatives are subject to compounded issues as expectations and requirements from all organizations involved create a complex regulatory landscape in which project specifications must satisfy concurrent and frequently competing demands. Last but not least, the resources—material, technical, and personnel—available at any given institution participating in such a research initiative only serve to complicate project management tasks by raising questions of fair contribution to project efforts based on performance obligation and credit received.

To better understand the implications of project management needs for digital medieval projects that bridge disciplines and institutions, we examine the experience of the Project Andvari initiative, funded by the National Endowment for the Humanities (NEH) Office of Digital Humanities, to highlight the costs and benefits of pursuing collaborative research in the face of discipline-specific project needs and the complicating factors of institutional demands.⁴

Shared Resources, Contributions, and Credit

Collaborative digital humanities initiatives represent a sea change for conventional research in the field of medieval studies inasmuch as they require team-based approaches to complex tasks that may extend beyond the boundaries of any particular team member's expertise. While this necessarily changes the scope of a research initiative and forces primary investigators (PIs) to adopt the mantle of team leader or project manager, interdisciplinary engagement can benefit initial project design and final products by establishing a network of expertise and institutional resources that are potentially greater in terms of scope and scale than could ever hope to be achieved by a lone researcher or a single institution. This requires that those seeking out partners for collaborative research projects appreciate the complex ecosystem of inter-institutional relationships by fully assessing resource availability (i.e. human, material, and financial resources) at each participating institution and weighing the level of resource allocation

³ Leon, "Project Management for Humanists."

⁴ Grant number HD-51640-13, grant period 7/1/2013–12/31/2014; and grant number HD-248511, grant period 8/1/2016–1/31/2018.

against the roles and contributions of individual researchers to the overall project goals. For digital initiatives, the challenge of establishing inter-institutional relationships is further complicated by introducing questions of providing and sustaining technical infrastructure such as servers, application support, developer time, and ongoing storage of project tools and data.

Distributed, collaborative project teams also reconfigure traditional approaches to scholarship by complicating notions of labour. Considering the varied level of contributions and institutional benefit gained by participating in a given project, the challenge posed to PIs is to manage initiatives in such a way as to fairly balance expectations of input from individual participants, while also ensuring that project deadlines and deliverables are met. The interdisciplinary nature of project teams raises further questions: What is the best way to organize a team to accommodate input from a variety of specializations, in particular those outside the primary scholarly discipline of medieval studies, while still maintaining a focus on the stated project goals? Student workers too must be weighed in this equation. Digital initiatives provide opportunities for mutual benefit of researchers and students, with the projects gaining worker hours and diverse perspectives, while student workers gain project-based learning experience and academic credentials. However, the distributed nature of such project work, along with different levels of resource allocation, raise the question of how to fairly assign credit to short-term or task-specific student workers. These considerations then press PIs to take a holistic view of project planning and governance, forcing them to consider from the outset of an initiative how to best prepare for the challenges of coordinating varied resources and personnel while amplifying interdisciplinary voices through conscious managerial practices and equitable labour practices.

Project Andvari: A Case Study

The Project Andvari initiative was established in 2012 as a collaborative endeavour involving The Catholic University of America (CUA), The University of Mississippi, the Institute for Advanced Technologies in the Humanities (IATH) at the University of Virginia (UVA), and partnering academic and cultural heritage institutions.⁵ It is not unique among digital medieval projects in terms of the complexities associated with management of multi-partner interdisciplinary projects. The team observed a need for a digital aggregation portal to facilitate research on the material culture of early medieval northern Europe and to serve as a clearing-house through which disparate collections of art and artefacts could be discovered and linked. The availability of library and museum resources, collections of artefacts, and records through web-based platforms has generated an unprecedented level of access, promoting global availability of material that had once been irretrievable to all but the physically present. While this accessibility is

⁵ For more information and continuous updates on the project, see Andvari: A Portal to the Visual World of Early Medieval Northern Europe.

heartening to all of us for whom limitless travel budgets are an unattainable fantasy, localized and idiosyncratic organization and classification approaches complicate the prospects for meta-analysis and cross-collection connectivity.

With these guiding principles, the project was established and the initial task of identifying the appropriate team members began. We as the project co-directors (Kopár and Wicker), representing backgrounds in medieval art and archaeology, literature, and historical linguistics, brought domain-specific research knowledge to the initiative, but our limited skills, competencies, and perspectives in information technologies and information science was acutely felt. In order to secure both the dedicated skills and requisite technological infrastructure needed to develop the proposed platform, we consulted and engaged the Institute for Advanced Technologies in the Humanities at the University of Virginia as a partnering organization. IATH brings a legacy of digital humanities innovation as well as the individual expertise of Daniel Pitti and Worthy Martin, experts in markup languages, database architecture and design, and digital perception systems operating in distributed environments. After a formalized agreement was established, a project proposal was developed in collaboration with the three hosting institutions (CUA, Mississippi, and UVA).

The Initial Steps of Project Management

Managing a new digital project starts with two essential steps after dreaming up a new initiative but before real work, and real project management in the traditional sense, can begin. The first step is assembling an effective team; the second, applying for grants to secure funding for project activities.

Assembling the project team

An interdisciplinary project, as all digital humanities (DH) projects in medieval studies are, requires a team with specialists in a variety of fields necessary to provide adequate support for each component of the project. These experts include, but are not limited to, authorities chosen for their knowledge of content in addition to specialists in technical matters and information management. Since most projects in digital medieval studies originate with scholars who have a specific research project (and research questions) in mind, content experts are not hard to find through traditional academic networks. The content experts of Project Andvari have long known each other through conferences and collaborative projects, and many of the initial discussions about the goals and desired platform functionalities of our project took place at academic meetings (conferences, workshops). The project was designed with an understanding that interdisciplinary engagement would benefit the project design as well as the final product. Accordingly, we identified further colleagues based on their domain-specific knowledge and invited them to collaborate. They included researchers from various scholarly fields (art history, archaeology, literature, linguistics, epigraphy, etc.) as well as institutional contributors (curators, museum experts, information specialists). Finding collaborators in the technical field was a bit more complex. Through our researchers, we identified technically

skilled colleagues who have already been involved in DH projects in related fields, but we needed to find a digital home and steady technical support for the project. Once we decided that the project would be based in the US, and confirmed that our own home institutions could not adequately support and sustain a digital initiative at the scale of our project, neither in terms of permanent technical staff, nor in infrastructure, we conducted an environmental study of institutional support for digital humanities projects at established centres of DH research. Based on data available online (at institutional and project websites) as well as through private communications, we explored the types of projects run by centres of DH scholarship primarily in the greater Washington, DC, area. Based on their research profile and project history, Daniel Pitti and Worthy Martin at IATH at UVA were identified as ideal collaborators for our project (see above). We conducted preliminary discussions via email and in person, and agreed to seek funding for an international collaborative project to open a dialogue between scholarly fields and experts with different technical capacities, spanning institutions in academe and cultural heritage preservation.

Grant writing

Internal institutional grants for faculty research, if available at all, can rarely be obtained for collaborative digital initiatives involving more than one institution. In addition, institutions are reluctant to pay sustaining costs (software, equipment, technical support, data storage, etc.) for digital humanities projects. It is necessary, therefore, to seek external funding through grants—normally through multiple grants. The landscape of support for digital projects is ever-changing, thankfully for the better, with a number of major and minor funding options available from start-up to implementation. Such grant opportunities include: numerous grants sponsored through the NEH Office of Digital Humanities (ODH),⁶ the Institute for Museum and Library Services Leadership grants,⁷ Europeana research grants,⁸ the Council for Library and Information Resources awards and fellowship offerings,⁹ and more.¹⁰ However, writing a collaborative grant is an art form that requires organizational (and often interpersonal) skills and a mastery of navigating diverse expectations and regulations of participating institutions.

The Project Andvari team wrote three major grant proposals in a period of three years, each requiring several months of organizing. Given the collaborative nature of the project, decisions had to be made about the host institution (weighing partner institutions of various levels of involvement and commitment) and about the nature of

6 National Endowment for the Humanities, “Grant Opportunities.”

7 Institute of Museum and Library Services, “Apply for a Grant.”

8 Europeana, “Europeana Research.”

9 Council for Library and Information Resources, “Awards and Fellowships.”

10 The University of Florida Center for the Humanities and the Public Sphere maintains a listing of granting institutions (www.humanities.ufl.edu/digitalhum-funding.html). *Digital Humanities Now* also regularly posts on newly available grants and fellowships. See Digital Humanities Now, “Funding and Opportunities.”

sharing costs and available resources. The initial grant application involved three universities: two public state universities (Mississippi and UVA), and one private institution (CUA), home institutions of the project co-directors and IATH. When selecting the host institution, we considered a number of practicalities: availability and willingness of the local project team member to serve as principal investigator (PI); availability of local resources (human, material, and financial) relevant for the proposed project activities; willingness of the institution to cost share; level of support provided for the PI (administrative help with grant management, release time, etc.); support of the grant office; and lastly, the benefits the project can bring to the institution (e.g. by providing training for graduate students and encouraging interdepartmental collaboration among faculty).

Phase I of the project centered on workshops and meetings of international collaborators; thus, CUA was selected as the initial host, based on its convenient geographical location in Washington, DC, providing easy access for international and East-Coast travellers, and its preparedness to assist with the proposed activities. CUA could also provide local conference organization, free access to the workshop venue, and administrative support. The grant money available was too limited to secure a course buy-out or release time for the PI; as a result, institutional support through graduate student volunteers and local administrators was vital for success. With the support of an NEH Digital Humanities Start-Up Grant (Level I), the project team convened the proposed international workshop on overall project design and—with a no-cost extension of the grant period—added a second workshop on the development of a project-specific controlled vocabulary on iconography, responding to a need identified during the first workshop.

A new cycle of grant writing began for Phase II of the project, which will include platform design, coding and development, and testing cycles. Due to the change in primary activities from workshops to technical tasks, we had to reconsider our choice of host institution. Since technical support for Project Andvari has been provided by IATH at UVA since the beginning of the project, we decided to designate UVA as the host institution for the next grant. The decision was again based on practical considerations: the willingness of project team members with Phase-II-specific technical skills to serve as PIs; existing local resources (human and technical) to support proposed grant activities; availability of local staff to assist with grant writing and administration; and ease of administering the grant (minimizing transfer of funds between institutions and the number of subcontracts needed). For Phase II, the other project co-directors are subcontracted through their home institutions. CUA agreed to cost-share and provide limited release time for Kopár to work on the project. Although The University of Mississippi cost-shared Wicker's time on Phase I of the project, they declined this arrangement for the second phase due to the concerns of new university administrators about the burden of reporting in light of the relatively small size of the grant (compared to NSF and NIH funding, for instance). Instead, a subcontract with the institution will provide funds to hire an adjunct instructor so that Wicker can be granted release from one course during the fall of 2017. To maximize our efforts, we also decided to hire a project manager, Joseph Koivisto, who had been involved in the project since 2013, first as graduate student volunteer and later as graduate research assistant provided to the

project by CUA as a sign of institutional support. His current institutional affiliation with the University of Maryland brings in a fourth participating institution—with new institutional regulations and expectations to further complicate the picture.

In addition to securing the support of the home institutions of the project co-directors and core team members, we had to acquire agreements from our partner institutions regarding data use and data management. For Phase II, our institutional partners for data sharing include three cultural heritage institutions in two different countries: The British Museum (with the Portable Antiquities Scheme) in London, UK; SOCH (Swedish Open Cultural Heritage) of the Swedish National Heritage Board, Stockholm and Visby, Sweden; and Norwich Castle Museum and Art Gallery, Norwich, UK. The agreements with these institutions also entail a human resources component to account for the time and efforts of our local project team members at these. We were pleased to find that these cultural heritage institutions show great interest and enthusiasm towards data sharing; thus, establishing formalized data agreements was an easy affair. However, the three institutions selected for the pilot project differ in their availability of local technical support and human resources. The British Museum and the Swedish National Heritage Board, both major national organizations, agreed to provide local technical support, while the much smaller Norwich Castle Museum would in part need to rely on our team. This creates challenges for project management to accommodate manual data transfers, provide greater project involvement in aligning data and ensuring quality of metadata records, and develop data sharing and usage agreements. However, these institutions were chosen for the pilot in part exactly because of the differences in institutional organization and availability of support. Going forward, Project Andvari plans to engage a variety of institutions as partners for data sharing, so it seemed wise to design a pilot not only to assess the feasibility of the proposed digital portal but also to test different levels of collaboration in project management. The project is currently in the early stages of Phase II, so the success of our plan remains to be seen.

Bridging the Gap between Disciplines

The researchers involved in Project Andvari—as well as many members of the scholarly community that we envision as beneficiaries of the eventual project—identify themselves as medievalists. Medieval studies is not a defined academic discipline, and scholars trained in or active in this field come from various academic traditions and are often concerned with issues that cross disciplinary boundaries. From the outset of the project, the Andvari team was constructed to promote interdisciplinary engagement with problems in medieval scholarship that temptingly dangle beyond the scope of a single scholar or a single discipline. Faculty members who carry out such research typically have “tenure homes” in specific disciplinary departments; for instance, Andvari project co-directors Kopár and Wicker are housed in a Department of English and a Department of Art and Art History, respectively. However, both are medievalists and have training in multiple fields, Kopár in literature, historical linguistics, and iconography, and Wicker in art history, archaeology, and Germanic philology with a PhD from an interdisciplinary program. Bridging the disciplines is a natural perspective for us.

As a digital humanities undertaking, Project Andvari is wider than medieval studies and brings together a team of researchers with various skills and knowledge, including information technology and data management. Thus the team collaboratively addresses a problem too broad for an individual or even a group of medievalists to tackle alone in an interdisciplinary (some would prefer to refer to it as multidisciplinary) research project. By including representatives from numerous disciplines, early discussions of the proposed project benefited from a multiplicity of perspectives, which brought to light complementary considerations of information-seeking behaviour, platform design, and scholarly practices within our identified audience. Ongoing cooperation with information technologists, various content experts (medievalists), museum professionals, and librarians helps to ensure that project planning, activities, and institutional collaborations are all considered through a diverse set of disciplinary perspectives. Digital humanities as a meta-perspective provides the glue that holds us together, building a “bridge between domains.”¹¹

Some disciplines lend themselves to collaborative, on-the-ground teamwork more than others. While linguists might work together on, for instance, a large, long-term dictionary project, the individual scholars would likely carry out their separate portions of the project from their own offices at institutions scattered across the nation and across the globe. In contrast, archaeologists typically accomplish at least part of their work physically engaged with each other excavating a site. While much can be accomplished with long-distance electronic communications via email, Dropbox, and Skype, face-to-face workshop interaction still offers unique opportunities to brainstorm and develop projects. We are grateful to the NEH for granting us support to bring an international group of scholars together for the initial Andvari workshop in 2013 and the smaller follow-up meeting in 2014. In the following section, we describe the planning and execution of these workshops as case studies, examining what worked and what did not work as well to advance the organization and team-building for this digital humanities project.

Lessons Learned: Management and Organizational Issues

The initial Andvari workshop proceeded very smoothly and established an enthusiastic tone of collaboration for the remainder of the project. The project team spent a great deal of time developing the workshop agenda in advance. We posted the preliminary agenda, together with proposed readings and links to relevant electronic resources, on our project blog.¹² The blog also presented logistical information about the workshop, and through its comment feature, allowed participants to contribute. Although the blog was initially designed for the workshop, we decided to maintain it and continuously post updates on our progress, as well as further links and readings.

Despite our successes, a project of this scope cannot be executed without frustrations. We, the project co-directors, were somewhat discouraged by the minimal

¹¹ Flanders, “Time, Labor, and ‘Alternate Careers,’” 297.

¹² Project Andvari: “Workshop Blog.”

preparation by some initial workshop attendees in reading their “homework” of articles and links provided on the project blog and also by their relative lack of input after the workshop. Attempts to solicit regular input from them regarding the development of project-specific thesaurus resources were not successful, hindering our attempts to generate initial work on our project-specific thesauri. Our sense is that the manner in which we approached workshop attendees for contributions—via emails and shared-access files—may have been an unproductive means of requesting input. We now realize that we needed to find ways to reconfigure our requests to encourage active participation from more of our geographically remote colleagues when we and the project are “out of sight, out of mind.” In fact, we co-directors are so busy with our own research and teaching that we did not devote a great deal of time to seeking ways to involve those who gave little or no input because it was more productive for us to move on. We are heartened that some workshop attendees still remain deeply engaged and offer continued support.

The second workshop was a bonus that was not planned for in the Phase I grant proposal. Compared to the initial workshop, this iconography thesaurus meeting was on a smaller scale, with limited international participation but with new US contributors. We did less extensive planning for it due to limited resources and time, and different levels of familiarity with the goals and history of the project among the participants caused some difficulties that were different from those encountered with the first workshop. We later realized that the additional workshop could have benefitted from more in-depth pre-workshop communication, particularly because we were bringing some new participants aboard for this project meeting. The general agreement was that those new to the project did not have the appropriate amount of foundational information about Project Andvari and therefore were not as prepared as desired for the workshop in terms of (1) decisions that had already been finalized; (2) the planned direction of the project after the workshop; and (3) technical information that will guide the project going forward (i.e. encoding protocols, envisioned platform, data providers, etc.). Due to this gap in understanding, time was spent bringing some attendees up to speed on the project as well as addressing questions and concerns that were peripheral to the conversation at hand. In the future, diligence should ensure that all attendees have access to appropriate documentation of previous project decisions and work plans. In addition, we should provide mandatory readings prior to the workshop to guarantee that attendees have the required level of understanding of content and technical concepts, and we must stress the importance of reviewing pre-workshop materials in order to save workshop time and maximize efficiency. Nonetheless, we were able to achieve meaningful progress in the second workshop toward the development of a project-specific iconographic vocabulary to describe early medieval northern European artefacts consistently. A substantial draft of the thesaurus has been created and uploaded into the Protégé application in a semantically structured format (see below).¹³

13 Stanford Center for Biomedical Informatics Research, Protégé.

How can project leaders require that academics with myriad responsibilities at far-flung institutions prepare in advance? The scholars must have incentives to become invested in the project, but it is difficult to engage newcomers in a venture that they do not fully understand, and it is equally difficult for them to understand a project that takes them beyond the comfort zone of their own disciplines. For instance, some of the archaeologists we brought to the initial workshop provided us with excellent ideas about the desired features of the Andvari portal from a research and user perspective, but did not have the technical knowledge to follow discussions of how the goal could become attainable and what the technological options and limitations were.

The vocabulary workshop in particular highlighted the gap in understanding between art historians and archaeologists on one side, and experts in information technologies and information science on the other; the gap between “know what” and “know how.”¹⁴ Some discussions were derailed by conversations springing from content experts’ limited understanding of the Simple Knowledge Organization System (SKOS)¹⁵ encoding protocol to be used for the ultimate product of the workshop, which in turn led to misunderstanding of how the initial word/concept list should be formatted. This required additional workshop time in order to explain the ways in which the protocol will accommodate the semantic structure of the vocabulary. We should not have assumed that technical concepts were equally familiar to all participants. As a remedy, we should have either provided additional introductory materials for attendees or dedicated a preliminary workshop presentation to explaining the characteristics of the SKOS protocol. In the future, as we bring new collaborators into the project, we plan to develop brief video tutorials to explain specific technical issues to scholars from various disciplines, along with introductory videos or tutorials for new users of Project Andvari once it is operational and available to the public.

As we develop our project, we also want to be mindful about bridging the gap within the overall project management. Team-building is crucial to a successful project, and communication within a team of geographically dispersed disciplinary scholars and digital humanities specialists must be more nuanced than a “reply-all” model focused on written communication with a sense of limited personal liability of the participating individuals.¹⁶ Face-to-face communication (in person or with the help of online conferencing tools, in small and focused groups or for the full team) is vital for discussion-intensive decision-making and feedback, as well as for developing a sense of community

14 Muller and Young, “Disciplines, Skills and the University,” 137.

15 The Simple Knowledge Organization System (SKOS) is a World Wide Web Consortium-recommended protocol that supports the semantic and relational data modelling needed to encode hierarchical authority records such as controlled vocabulary, subject headings, and thesauri. Used in conjunction with the Resource Description Framework (RDF) protocol, SKOS allows for concepts to be linked via semantic (broadMatch, closeMatch, exactMatch) and hierarchical (broader, narrower, inScheme) elements, allowing for complex linkages to span concepts within a given scheme and across third-party authorities, thereby supporting linked open data best practices. Introductions to SKOS can be found at W3C, “Introduction to SKOS.”

16 Siemens, “It’s a Team If You Use ‘Reply All,’” 225.

behind and beyond a digital tool, which promotes buy-in, project ownership, and sustainability. Further, it is important to consider continuous involvement of new team members, especially short-term and long-term volunteers, to ensure the sustainability, flexibility, and growth of the project over time. Engaging the next generation of digital humanities experts and practitioners, our graduate students, is a major step toward achieving this goal, and much more.

Graduate Students as Agents of Interdisciplinarity

During the early stages of the initial grant period, it was determined that Project Andvari could serve as an experiential digital humanities training opportunity for graduate students. The addition of a graduate research assistant would provide us with a general aide for day-to-day administrative functions and also extend the project's function beyond its primary goals by supporting the development of interdisciplinary and collaborative research skills in the next generation of researchers. This type of experience would also provide participating students with market-ready experience and credentials that would aid in their post-graduate career search. Such project-based learning as practiced in the CEDAR project¹⁷ and at digital humanities centres such as the Maryland Institute for Technologies in the Humanities represents current trends in digital humanities education.¹⁸

In conjunction with the Department of Library and Information Science (LIS) at CUA, project co-director and CUA faculty member Lilla Kopár identified and invited LIS graduate student Joseph Koivisto to work part-time as a research assistant. After participating in the initial project workshop, he expressed interest in working with the project team as it would provide a useful enhancement to his graduate coursework focusing on the management of cultural heritage collections in so-called GLAM institutions (galleries, libraries, archives, and museums). Funding for a part-time research assistant was secured through generous support from the Dean of the School of Arts and Sciences, ensuring that student labour would be compensated—as will be discussed below. The inclusion of LIS graduate student perspectives was valuable to us as the insight on LIS-specific disciplines such as information-seeking practices, controlled vocabulary and thesaurus development, and human-computer interactions (all topics covered in the core curriculum of CUA's LIS graduate program) helped to further refine the proposed system development to promote usability. Furthermore, the technology-based skills now frequently taught within LIS programs—database and web-coding, metadata, and linked open data (LOD)—had properly prepared Koivisto to engage with technical subject experts from IATH and other institutions. Despite having no background in medieval studies, his technical capabilities allowed him to quickly adapt to

¹⁷ Ensslin and Slocombe, "Training Humanities Doctoral Students." This article addresses the constructivist nature of project-based, collaborative learning described by Florian Schneider, "Collaboration."

¹⁸ Fraistat, "The Function of Digital Humanities Centers," 281.

project-specific interlanguage of information systems specialists and traditional humanities researchers.¹⁹

While libraries and library science graduates have been called out in particular for their value to and integration with digital humanities initiatives,²⁰ the participation of graduate students in any DH project—both from within and without the domain of the primary investigators or project directors—represents an opportunity to expand the critical perspective of the involved researchers, increasing the chance that the resultant platform or product will extend beyond the traditional confines of domain silos.²¹ Digital humanities initiatives are upheld as interdisciplinary at their core due to the frequently varied domain backgrounds of collaborators. Our project is no different; the project team and workshop attendees represented fields including medieval history, art history, archaeology, literature, information technology, system design, and more. However, the introduction of graduate students alters the team dynamic by providing unique domain-specific perspectives that are free of the encumbrances of career-long participation—and indoctrination—in the conventions of a given scholarly tradition. Distinctly positioned to occupy the emerging interzone between traditional knowledge work and technological applications—the “know what” and the “know how”²²—graduate students of all domains have been exposed from an early stage of their education to a model of scholarly work and research that habitually extends beyond the borders and boundaries of traditional discipline and practice. Considering the influence that faculty advisors and project supervisors play in the formation of future research patterns,²³ the inclusion of graduate students of diverse disciplinary backgrounds in DH initiatives sets the stage for the realization of several benefits to the next generation of faculty and researchers: examples of collaborative, interdisciplinary engagement; experiential learning of applied domain knowledge in conjunction with project management skills; and exposure to the digital skills, tools, and knowledge that are increasingly important to all agents of the digital humanities.

Additional graduate students have engaged with the project through voluntary contributions to project workshops and ongoing project support. Beth Newman Ooi, a doctoral candidate in the Medieval and Byzantine Studies Program at CUA, assisted with planning and execution of the initial project workshop in 2013. This project aligned in content and geographical scope with her own research on the material culture of the early medieval period, allowed her to gain first-hand experience in the planning and execution of an interdisciplinary scholarly workshop, and last but not least, provided her with an opportunity to meet international scholars in her own field of study. Additional library science graduate students contributed to the project through voluntary

¹⁹ Svensson, “The Digital Humanities as a Humanities Project,” 58.

²⁰ Gailey and Porter, “Credential Creep in the Digital Humanities”; Bryson et al., “SPEC Kit 326: Digital Humanities.”

²¹ Cromity and de Stricker, “Silo Persistence: It’s Not the Technology.”

²² Muller and Young, “Disciplines, Skills and the University.”

²³ Barrett, “The Information-seeking Habits.”

coursework in collaboration with Kevin Gunn, subject librarian and lecturer in the LIS Department at CUA, who taught a special topics course on digital humanities within the context of academic research libraries. These students provided much-needed support to the project through several avenues. Kirsten Mentzer, Katie Kane, and Jeremy Withnell completed a survey of existing controlled vocabularies and subject headings in order to assess the availability of subject-specific descriptive terminologies for the project's identified focus. Anita Kinney developed formalized financial reports for the project, thereby supplementing our record keeping for final reporting to the NEH. Samuel Russell investigated alternative funding opportunities for the project, compiling a comprehensive list of regional, national, and international funding bodies to whom we could apply. Their work proved invaluable help to Project Andvari. More importantly, however, the experience of working on an interdisciplinary initiative under the close supervision of their professors provided the students with a project-based educational opportunity that helped to bridge the divide between DH theory and practice.

A Note on "Fair Trade"

The Project Andvari leadership views the interdisciplinary makeup of its team as one of its enduring strengths. However, the prospect of bringing in a diverse group of project members—especially with regards to graduate student workers and volunteers—is not a neutral element. Volunteer labour, in particular that associated with classroom instruction or coursework, represents a huge resource pool for projects; however, as Spencer Keralis notes, such labour represents a model of deficit internship that shifts project costs to the volunteer students.²⁴ Furthermore, the interdisciplinary nature of much DH work can create relationships that reinforce harmful power dynamics between traditionally hierarchical roles: faculty to graduate student workers, the professoriate to librarians, and scholarly researchers to project volunteers. Such dynamics not only position the faculty researcher as the centre of an otherwise collaborative and innovative endeavour,²⁵ but also re-establish the internalized service nature of those viewed as primarily outside the privileged position of professional scholar, namely librarians, information technologists, graduate and undergraduate students, and community volunteers.²⁶

Project-based experience and the application of coursework concepts to project work still retain value, but it is important to recognize that investing in future DH researchers requires equitable labour practices, which include remuneration and recognition. Even short-term assistantships can provide students with some financial support as well as a formal tie to the project, a link that can prove advantageous when applying for employment both within and without the university. Further, while the immediate candidate pool for graduate assistants is often within the faculty member's home department, it

24 Keralis, "Milking the Deficit Internship."

25 Muñoz, "Digital Humanities in the Library."

26 Nowviskie, "On Capacity and Care."

is important to try to extend research assistantship positions to other departments as well. This increases the diversity of disciplinary perspectives and may serve as a bridge towards future interdepartmental collaboration. While it may seem that the pursuit of equitable labour values and disciplinary diversity regarding student workers places an additional burden on the project directors, it does not have to be an oversized task. Simple approaches to this issue include:

- Review the “Collaborators’ Bill of Rights”²⁷ and the “Student Collaborators’ Bill of Rights”²⁸ to ensure familiarity with the rights of student workers and the expectations placed upon faculty researchers and investigators. Pay special attention to the distinctions drawn between paid and unpaid student internships.
- Speak with department heads or deans regarding their willingness to provide funding support for graduate assistantships for students outside of the immediate department or school.
- Meet with graduate advisors and mentors from programs such as computer science, library and information science, geographical sciences, and other non-traditional avenues of partnership to determine their interest in or ability to support graduate student collaboration.

Lastly, formalized recognition and support at the project, departmental, and institutional level should support student production of scholarly output based on project participation. Julia Flanders highlights the importance of “devoting resources to creating meaningful job and internship opportunities at digital humanities research projects, scholarly publications, conferences and other professional activities.”²⁹ In addition to reaffirming the value of graduate student involvement, this approach will enhance the project-based experience of graduate assistants by allowing them to produce novel input to the ongoing digital humanities discourse and will promote project visibility by using multiple outlets, potentially across several disciplines. It will also provide graduate assistants with a means to secure credentials that will prove important as they search for permanent posts with universities and GLAM institutions. Based on project activities, Koivisto developed and delivered several conference posters and presentations at CUA workshops, symposia, and conferences including the Society of American Archivists (SAA) annual meeting and a special conference at the German Historical Institute on the implementation of crowdsourcing approaches for the creation of historic knowledge.³⁰ One conference poster evolved into a peer-reviewed publication through the SAA’s annual research forum.³¹ These presentation and publication opportunities not only afforded our graduate assistant the chance to obtain refereed publication—an

27 Kirschenbaum et al., “Collaborators’ Bill of Rights.”

28 Di Pressi et al., “A Student Collaborators’ Bill of Rights.”

29 Flanders, “Time, Labor, and ‘Alternate Careers,’ ” 307.

30 German Historical Institute, “Creating Historical Knowledge Socially.”

31 Choi and Koivisto, “Controlled Vocabulary Enhancement through Crowdsourcing.”

important element of any academic curriculum vitae—but also extended project exposure to the domain of archival studies, promoting further project awareness and future collaborative potential outside the sphere of conventional medieval studies.

Project Management: A Neglected But Necessary Skill Set

The topic of project management in the context of collaborative humanities scholarship is one marked by tension. Although vital to the success of project-centred research initiatives, the role of the primary investigating scholar (PI) as an administrative or managerial functionary is viewed with a degree of scepticism due to the conventional division of scholarly and administrative/managerial work. The former is understood to be the “knowledge work” traditionally associated with academia, while the latter constitutes the quantifiable, productivity-driven work of those considered outside the scope of pure scholarship.³² The situation is further complicated by the fact that digital projects also require experts in technical fields with skill sets seldom mastered by the discipline-based content experts. In the current climate of academic institutions, the “know what”—conceptual knowledge that makes up the narrowly defined domain knowledge of advanced scholarship—stands firm as the primary responsibility of the academy. In contrast, the “know how”—applied technical concepts that include coding, system design, and content management in addition to project management—are, at best, secondary considerations or, at worst, harbingers of the “de-specialization” of the university.³³ The positioning of disciplinary knowledge and managerial knowledge as antithetical areas of expertise gives rise to professional scholars who find themselves unprepared for project management responsibilities.³⁴

For digital medievalists, the increased prominence of collaborative interdisciplinary initiatives has marked a global shift in the approaches and methodologies needed to facilitate successful research within the narrow domain of medieval studies. Yet current instructional offerings and professional development opportunities for scholars may be insufficient to prepare them adequately for project management functions such as supervising interdisciplinary project teams and handling institutional expectations for inter-institutional initiatives. At the “Digital Approaches in Medieval and Renaissance Studies” workshop held at the University of Oxford in July 2015—an event whose stated goal was “to foster collaborative working, creative thinking, and user-focused project development”—project management was not featured in any session during the five-day workshop.³⁵ This is somewhat surprising given the acknowledged importance of project management skills to scholars increasingly tasked with self-managing initiatives that

32 Flanders, “Time, Labor, and ‘Alternate Careers,’” 293.

33 Muller and Young, “Disciplines, Skills and the University,” 137–138.

34 Leon, “Project Management for Humanists.”

35 For more information and the agenda, see Sieftring, “Digital Approaches in Medieval and Renaissance Studies.”

cross both scholarly and technical disciplines, as attested by workshop organizer Judith Siefring (Bodleian Libraries, University of Oxford).³⁶

The majority of publications, resources, and educational opportunities that focus on project management aim primarily at promoting familiarity of and fluency with technologies and applications most frequently associated with digitally enabled research. Knowledge of databases, web applications, coding, and metadata is of great importance for medievalists inasmuch as it is a prerequisite of effective communication with system specialists and information scientists, yet it represents only a part of the overall project management spectrum. More importantly, medievalists tasked with steering an interdisciplinary project group should be capable of performing effective management in the form of resource allocation, timeline scheduling and management, and task delegation. Key to this realization is the understanding that the field of digital humanities is—by its very nature—a conceptual space of intersecting specializations. In this arena, numerous specialists and experts will make demands on time, resources, and consideration, all things that are owed to them but that must be deliberately governed in order to ensure appropriate distribution. As primary investigators, it is important to conceive of project management as a set of soft skills that will facilitate project growth and development. By consciously delegating tasks, managing work plans, and aligning expectations of all involved specialists, time, effort, and grant funds can be put to their best possible use.

Concluding Remarks

An interdisciplinary project is a source of both challenges and unexpected benefits; it is a means of continuous exploration and learning, sometimes through failures. The latter can, however, be minimized by efficient planning and management. Not only do conscientious planning and project stewardship create a sense of shared ownership among all project participants, but they establish a research dynamic that values input from a diverse range of contributors including technologists, interdisciplinary scholars, and students. We hope that sharing our experiences in the field of project management will provide new teams of digital humanists with ideas and valuable insights. As DH projects become increasingly collaborative with each other, it is valuable to find means of sharing not only our data and technical resources, but also our experiences of project management, without which no project can ever exist.

As Project Andvari continues to explore the methodological and technical requirements of generating an innovative platform for aggregated research, the project team strives to remain true to the foundational principles of our initiative: allowing users to study the visual material culture of the northern periphery of early medieval Europe across media and beyond traditional geographical and disciplinary boundaries. In this vein, we will continue to cultivate lasting collaborative relationships with contributing scholars and institutions, seeking their input throughout future development

³⁶ Personal correspondence, October 26, 2015.

stages. Thus, we hope to bridge a gap that scholars have identified, providing a functional digital resource that will facilitate interdisciplinary research across collections, institutions, and nations.

Bibliography

- Andvari: A Portal to the Visual World of Early Medieval Northern Europe. Project website. www.andvari.org.
- Barrett, Andy. "The Information-seeking Habits of Graduate Student Researchers in the Humanities." *The Journal of Academic Librarianship* 31 (2005): 324–31.
- Bryson, Tim, Mariam Posner, Alain St. Pierre, and Stewart Varner. "SPEC Kit 326: Digital Humanities." Washington, DC: Association of Research Libraries. Last modified November 2011. <http://publications.arl.org/Digital-Humanities-SPEC-Kit-326/>.
- Choi, Youngok, and Joseph Koivisto. "Controlled Vocabulary Enhancement through Crowdsourcing: Project Andvari, Micropasts, and Public Quality Assurance." Society of American Archivists 2015 SAA Research Forum. Last modified July 2016. www2.archivists.org/sites/all/files/ChoiKoivisto-ResearchForumPaper2015.pdf.
- Council for Library and Information Resources. "Awards and Fellowships." www.clir.org/fellowships.
- Cromity, Jamal, and Ulla de Stricker. "Silo Persistence: It's Not the Technology, It's the Culture!" *New Review of Information Networking* 16 (2011): 167–84. doi:10.1080/13614576.2011.619924.
- Digital Humanities Now. "Funding and Opportunities." <http://digitalhumanitiesnow.org/category/news/funding/>.
- Digital Medievalist. "Project management." Last modified 2012. https://web.archive.org/web/20141001054235/www.digitalmedievalist.org/wiki/index.php/Project_management.
- Di Pressi, Haley, Stephanie Gorman, Miriam Posner, Raphael Sasayama, and Tori Schmitt. "A Student Collaborators' Bill of Rights." UCLA Digital Humanities. Last modified 2015. www.cdih.ucla.edu/news-events/a-student-collaborators-bill-of-rights/.
- Ensslin, Astrid, and Will Slocombe. "Training Humanities Doctoral Students in Collaborative and Digital Multimedia." *Arts and Humanities in Higher Education* 11 (2011): 140–56.
- Europeana. "Europeana Research." <http://research.europeana.eu/>.
- Flanders, Julia. "Time, Labor, and 'Alternate Careers' in Digital Humanities Knowledge Work." In *Debates in the Digital Humanities*, edited by M. Gold, 292–308. Minneapolis: University of Minnesota Press, 2012.
- Fraistat, Neil. "The Function of Digital Humanities Centers at the Present Time." In *Debates in the Digital Humanities*, edited by M. Gold, 281–91. Minneapolis: University of Minnesota Press, 2012.
- Gailey, Amanda, and Dot Porter. "Credential Creep in the Digital Humanities." *#alt-academy: A Media Commons Project*. Last modified May 6, 2011. <http://mediacommons.futureofthebook.org/alt-ac/pieces/credential-creep-digital-humanities>.

- German Historical Institute. "Creating Historical Knowledge Socially: New Approaches, Opportunities and Epistemological Implications of Undertaking Research with Citizen Scholars." October 26–28, 2017. www.ghi-dc.org/events-conferences/event-history/2017/conferences/2nd-ghi-conference-on-digital-humanities-and-digital-history.html?L=0
- Institute of Museum and Library Services. "Apply for a Grant." www.imls.gov/grants/apply-grant/available-grants.
- Keralis, Spencer D. C. "Milking the Deficit Internship." *Disrupting the Digital Humanities*. Last modified January 6, 2016. www.disruptingdh.com/milking-the-deficit-internship/.
- Kirschenbaum, Matthew G., Bethany Nowviskie, Doug Reside, and Tom Scheinfeldt. "Collaborators' Bill of Rights." Last modified January 21, 2011. <http://mcpress.media-commons.org/offthetracks/>.
- Leon, Sharon M. "Project Management for Humanists: Preparing Future Primary Investigators." Last modified May 6, 2011. <http://mediacommons.futureofthebook.org/alt-ac/pieces/project-management-humanists>.
- Muller, Johan, and Michael Young. "Disciplines, Skills and the University." *Higher Education* 67 (2014): 127–40.
- Muñoz, Trevor. "Digital Humanities in the Library Isn't a Service." Last modified August 19, 2012. <http://trevormunoz.com/notebook/2012/08/19/doing-dh-in-the-library.html>.
- National Endowment for the Humanities. "Grant Opportunities," www.neh.gov/grants/ODH.
- Nowviskie, Bethany. "On Capacity and Care." Keynote Address, NEH Office of Digital Humanities Project Directors Meeting, September 25, 2015. Last modified January 15, 2016. www.youtube.com/watch?v=LoGr3rPiGE8.
- ProjectAndvari: Workshop Blog. Last updated September 6, 2016. <https://projectandvari.wordpress.com/>.
- Schneider, Florian. "Collaboration," Summit: Non-Aligned Initiatives in Education Culture (blog), February 2, 2007, <https://web.archive.org/web/20141224141111/http://summit.kein.org/node/190>.
- Siefring, Judith. "Digital Approaches in Medieval and Renaissance Studies." University of Oxford, Digital Humanities at Oxford Summer School 2015. <http://dhoxss.humanities.ox.ac.uk/2015/digmedren.html>.
- Siemens, Lynne. "It's a Team If You Use 'Reply All': An Exploration of Research Teams in Digital Humanities Environments." *Literary and Linguistic Computing* 24 (2009): 225–33.
- Stanford Center for Biomedical Informatics Research. Protégé. <https://protege.stanford.edu/>.
- Svensson, Patrik. "The Digital Humanities as a Humanities Project." *Arts and Humanities in Higher Education* 11 (2001): 42–60. doi:10.1177/1474022211427367.
- W3C (World Wide Web Consortium). "Introduction to SKOS." Last modified January 1, 2012. www.w3.org/2004/02/skos/intro.

INDEX

- ABC Harmony (ontology), 212–13, 215
 Actor-Network theory, 94, 100, 103–4, 107
 Adam (Biblical figure), 54–56, 59–61
 Adams, Joseph Quincy, 73, 86
 Agas Map, 48, 87
 allusion, 45, 70, 125, 139
 American Association of Museums, 208
 Andrew W. Mellon Foundation, 123, 209
 Anglo-American Cataloguing Rules, 205
Annunciation (by Jan van Eyck), 208
 antifraternal literature, 75, 78
 ArcGIS (software), 47, 64
 archaeology, 82, 94, 226, 229, 234
 architecture, 4, 8, 10, 95
 Arts and Humanities Research Council (AHRC), 125, 139–40
 Asia, 54, 63
 auction catalogues, 205, 208–9
auctoritas, 123, 143–46
 authorial intention, 132, 136–41, 143–44, 155
 authority. *See auctoritas*
- Baltimore, 129–30
 Basilica di Santa Croce, 218
 Beatty, Alfred Chester, 218
 Bédier, Joseph, 124, 136, 147, 153
Beowulf, 4, 18, 28–36
 Berlin, 209
 Bernard of Chartres, 143
 Berry, David, 2
 Biblioteca Nazionale Centrale di Roma, 218
 Bibliothèque nationale de France fr. 1584 (MS A), 127, 132–33
 Bibliothèque nationale de France fr. 1586 (MS C), 125, 127, 133
 bildungsroman, 160
 black box, 93–94, 96, 100, 103, 109, 111
 Blackfriars, 5, 69–76, 79, 84, 86–89
- Blickling Homilies*, 4, 28–34, 36
 Bloch, R. Howard, 1
 Bodleian Library, 49, 144, 165–66, 210, 238
 Bonaventure (Saint), 80–81
 book history, 100, 194, 197
 Boolean operators, 125
 bootcamp (coding), 96, 107
 British Library Harley, 78, 75
 British Library Royal 2 B VII, 105
 British Library Royal 17 C. XXXVII, 5, 44, 48, 50, 52, 54
 British Library, 105–6, 165–66, 210
 British Museum, 8, 204, 229
 Bruni (MS buyer for Czar Nicholas I of Russia), 209
 Brussels, 209
 Bruzelius, Caroline, 79, 82
 Burrows, John, 42
 Busby, Keith, 124, 148, 150
- Calabrese, Michael, 154
 Cambridge University Library, 166, 197, 210
Canterbury Tales, 4, 28, 36–37, 39–40, 42, 146
 Carnegie Museum of Art, 209
 Carto (software), 47, 64
Catalogue of Manuscripts Containing Anglo-Saxon, 4, 13
 Cerquiglini, Bernard, 6, 145–55
chansonnières, 124, 138, 140
 Chartreuse de Champmol, 209
 Chaucer, Geoffrey, 36, 39–42, 78, 154, 157, 168, 179, 191, 195–96
 Cheltenham, 218
 Christ, 50, 52, 56, 60, 62, 80–81
 Chun, Wendy Hui Kyong, 95–96, 107, 109, 118
 citation, 125, 128–29, 136, 139
 Clapham, Sir Alfred, 71–74, 76, 81, 86

- Cleopatra, the birth of, 211
- close reading, 3, 27, 29, 34–36, 40, 42, 61, 64, 139, 188
- code mills, 96
- code, 2, 3, 27, 95, 108, 115, 128, 168, 185–86
- codicology, 100, 160, 164, 166, 180, 184, 196
- collaboration, 8, 28, 32, 109, 175, 188, 223, 226, 228–30, 233, 235–36
- Collaborators' Bill of Rights, 236, 239–40
- inter-institutional relationships, 223–25, 237
- limitations, 2, 27, 32, 35, 69, 95, 100, 110, 151, 155, 163, 196, 210, 232
- Comité International pour la Documentation Conseil International des Musées (CIDOC), 7, 210
- Conceptual Reference Model (CIDOC-CRM), 7–8, 210–13, 215, 219
- comma-separated values (CSV), 219
- computer-assisted analysis, 27, 35, 42
- cluster analysis, 4, 18–21, 28–33, 36, 39
- content, 6, 13–14, 17, 22, 32, 42, 48, 50, 63, 93, 96, 100, 103, 106, 109–11, 113–14, 119, 123–24, 130, 158, 167, 171, 173, 175–76, 179, 184, 188, 193, 213, 226, 231–32, 234
- experts, 8, 194, 226, 230, 232–33, 237
- controlled vocabulary, 228, 232–33
- cosmographia*, 48
- Council of European Research Libraries (CERL), 207
- critical apparatus, 3, 124–25, 129, 144, 151–53, 155, 175
- critical code studies, 2
- Cross (of Christ), 56, 59–62, 72
- crosswalks (metadata, data), 4, 7, 157, 177–78, 180–81, 185, 209
- and Digital Manuscript Index, 7, 159
- and e-Codices, 160, 175–78, 192–93, 195
- culture, 1, 7, 9, 46, 63, 75, 93–96, 98, 100, 110, 118–20, 123, 138–39, 142, 144–45, 148, 155–56, 160, 192, 195–97, 199, 204, 225, 234, 238–40
- analogue, 1–3, 7, 94–95, 97, 99, 118, 158–59
- cultural ethos, 95
- cultural heritage, 6–8, 108–110, 119, 158, 169–70, 173, 203–5, 210–11, 219–20, 225, 227, 229, 233
- and materiality, 2, 9–10, 100, 102, 135, 158
- Cuvier, Georges, 147, 155
- Da Vinci Code*, 212
- data curation, 159, 169, 191, 195
- data management, 223, 229–30
- data mining, 35
- data modelling and models, 4, 7–8, 205, 209–12, 215–16, 218–19, 232
- database, 64, 109, 111–12, 117–18, 158, 164, 205, 207–10, 216, 219, 226, 233, 238
- versus spreadsheet, 109
- deformance, 46, 62
- Deluz, Christiane, 5, 50, 52
- dendrogram, 4, 19–23, 29–37
- fliegende Holländer, Der*, 212
- Di Bacco, Giuliano, 125
- digital archive, 46, 51–52, 78, 110, 123, 128, 141, 161
- digital humanities, 1–4, 7–9, 35, 69, 78, 89, 95, 100, 107–8, 110, 160–61, 169–71, 175, 188, 190, 194–96, 223–24, 226–28, 230, 232–36, 238
- typologies of, 2
- Digital Index of Middle English Verse, 111
- Digital Manuscript Index (DMS-Index), 7, 159–60, 166–68, 170, 175–76, 179, 184–85, 188–89, 191–96
- digital mapping, 47–48, 50–51, 64
- digital materiality, 2
- Digital Walters, 169
- Digitally Encoded Census Information and Mapping Archive (DECIMA), 47
- digitization, 4, 8, 94, 99–100, 106, 157–60, 162–64, 166–67, 169–71, 173–74, 186, 188, 192, 195–97, 210, 213
- and ethics, 179, 196
- Dijon, 209
- display layer, 93
- documentation, 28, 32, 184, 205, 231

- domain silos, 234
 domain-specific approaches, 223
 domain-specific knowledge, 226
 domain-specific perspectives, 234
 Dominic, Saint, 78, 81
 Dominican Order (Order of Preachers),
 5, 69, 71, 73–75, 81–82, 86–87
 Doubleday (publisher), 212
 Doyle, Ian, 75–76
 Drout, Michael D. C., 4, 11, 13, 30, 33
 Drucker, Johanna, 35, 46, 175
 Dublin Core (DC), 51–52, 170, 172–78,
 192, 213

 Early English Text Society, 70, 144
 Earthly Paradise, 5, 45, 48, 51, 53–54,
 57–61, 64, 71
 Eden, 45
 Egypt, 61
 ekphrasis, 75–76, 90
 empiricism, 5, 78
 Europe, 193, 203, 216, 225, 238
 European Union, 204
 Europeana Data Model, 212–13, 220
 Europeana, 7, 212–13, 220, 227, 239
 event ontology, 215, 221
 Excel, 219
 eXtensible Markup Language (XML),
 6–7, 114–16, 120, 125–26, 142, 153,
 169–71, 173–74, 177, 180, 185, 187,
 190, 198
 eXtensible Stylesheet Language
 (XSL), 115–16
 transformations (XSLT), 129
 and WordprocessingML, 115

 failure, 28, 34, 42, 54, 59–60, 77, 123–24,
 140, 179, 191, 223, 238
 fair trade, 235
 Fenwick, Thomas Fitzroy, 218
 file system, 97
 folderless file system, 98
 Macintosh File System, 97, 119
 file/folder metaphor, 6, 94, 96–98, 107,
 118, 120
 Flanders, Julia, 128–29, 140–41, 230,
 236–37, 239
 Florence, 47
 formes-fixes, 124
 formula density analysis, 29, 34, 36
 fourteenth century, 5–6, 45, 69–70, 78,
 80, 123–24, 132, 139, 144
 Foys, Martin, 47–48, 66, 164, 199
 Francis, Saint, 78, 80–81, 90
 Franciscan Order (Order of Friars
 Minor), 5, 69–71, 81–83, 86–87, 91,
 166, 206
 French of Italy, 47
 Friedman, John, B., 48, 66
 function words, vii, 13–15, 31–32, 37–39
 Functional Requirements for
 Bibliographic Records (FRBR), 7–8,
 207, 211–13, 216
 funding, 79, 94, 139, 153, 159, 187, 223,
 226, 227–29, 233, 235–36

 Gärdenfors, Peter, 215
*General Prologue (Canterbury
 Tales)*, 36–39
 geo-temporal exhibits, 5, 51, 64
 Geoffrey of Vinsauf, 194
 Geographic Information System (GIS)
 mapping, 1, 47, 64
 Gerald of Wales, 194
 German Historical Institute, 236
 Getty Research Institute, 236
 Gillespie, Vincent, 149
 Galleries, Libraries, Archives, and
 Museums (GLAM) institutions, 3, 8,
 233, 236
 Golgotha, 61
 graduate students, 94, 163, 223,
 228–29, 233–36
 grant writing. *See* funding
 Gray, Freddie, 130
 Great Chan, 56
 Greek Orthodox Scriptures, 59–60
 Greyfriars, 79, 82–86, 88
 Guglielmo Libri, collection of, 216–17

 Hanna, Ralph, 144, 154
 Hassall, T. G., 82–83
*Helas! pour quoy virent onques mi oueil
 (Lo53)*, 125–27
 Henry VIII, King, 78
 Herculaneum, destruction of, 211

- Hereford mappa mundi, 5, 45–46, 50–65
 Higgins, Iain Macleod, 48, 50–51,
 54, 58, 60
 Hoccleve, Thomas, 190–91, 195–96
 HOLLIS (library catalogue), 7, 207
 Holsinger, Bruce, 75–76, 165
 Holy Trinity, Long Melford, 110, 117–18
 HopkinsLocal, 130
 Hudson, Anne, 149
 humanities scholarship, 1, 2, 4, 9,
 34–35, 42, 46, 63, 78, 106, 108, 170,
 223, 237
 Imperial Hermitage Museum, 209
 incipit, 125, 136, 181, 184, 189
 India, 54
 information storage, 97, 98, 107,
 225, 227
 informational palimpsest, 110
 inscription, 62–63, 96, 103, 109–11, 205
 Institute for Advanced Technologies
 in the Humanities (IATH), 8,
 225–28, 233
 Institute for Museum and Library
 Services (IMLS), 227
 institutional support, 225, 22–29,
 233, 236
 intentionality. *See* [authorial intention](#),
 [scribal intention](#)
 interdisciplinary, 223–26, 229–30,
 233–35, 237–39
 interoperability, 51, 93, 95, 161, 163–64,
 166–67, 170–71, 173, 186
 intertextuality, 138, 139
 intratextuality, 133n, 136
 Isidore of Seville, 48, 189
Itinerarium Kambriae (by Gerald of
 Wales), 51
 Jauss, Robert, 134
 JavaScript Object Notation (JSON), 219
Je Chante Ung Chant, 6, 123,
 124–29, 138–40
 Jerusalem, 45, 48, 51–52, 54n, 56,
 60, 62–63
 John le Long, 64
 Johns Hopkins University, 123, 130
 Kane, George, 77, 89, 149–50
 Ker, Neil, 4, 13
 Keralis, Spencer, 235
 Koh, Adeline, 2, 161
 Koszycki, Michal, 5, 69, 78
 Kringla (database), 8
 Kringla Leaf, 211
 Lab1100 group, 216
 Lachmann, Karl, 124, 136, 153
 Langland, William, 69–70, 75, 77–78, 80,
 86, 151, 154
 Last Judgement, 60–61
 Lawton, David, 76, 144, 154, 180
 Levenshtein distance, 125
 lexeme, 31
 Lexomics, 4, 11–13, 18, 19n, 28,
 30–33, 36–42
 Lexos (software), 29, 32
 library and information science (LIS),
 223, 233, 236
 Library of Congress, 165, 167, 170,
 173, 212
 Lightweight Information Describing
 Objects (LIDO), 209, 211, 219
 linked open data (LOD), 219,
 232n, 233
 Linking Open Descriptions of
 Events (LODE), 215
Livre de merveilles, 5, 52
Livre dou Voir Dit (Voir Dit),
 133n, 134–35
Loanges des Dames, 125
 London, 5, 48, 69–71, 73–74, 76, 86–87,
 165, 216–18, 229
 Lydgate, John, 6, 94–95, 108–118, 159,
 162, 180, 191–92, 195–96
 lyric (genre), 123, 124–25, 128–29, 134,
 136, 139–40, 179, 191
 M. Knoedler & Co, 209
 Machan, Tim William, 149, 153
 Machaut, Guillaume de, 6,
 123–25, 132–41
 complete works, 124, 132–33
 Manuscript A. *See* *Bibliothèque
 nationale de France* fr. 1584

- Manuscript C. *See* *Bibliothèque nationale de France* fr. 1586
- Machine-Readable Cataloguing (MARC), 7, 173–74, 205, 207
- Mandeville, John, 5, 45–46, 48–62, 64–65
- Manuscript, 3–6, 8, 11–12, 49, 50n, 52, 56, 63–64, 70, 75, 94–95, 99, 102–3, 107, 110, 115, 123, 134–37, 139–41, 150–51, 154, 164–65, 169, 191, 203–5, 207, 211, 215, 218–19
- and base manuscripts, 125, 129, 132, 138, 144
- and books of hours, 100
- and Codex Gigas, 100
- and complete works' manuscript(s), 124, 132–33
- and digitization, 8, 94, 99, 105–6, 157–59, 160, 162–64, 166, 167, 170–71, 173–74, 186, 188, 192, 195–97, 210
- and extant manuscripts, 34, 63, 124, 125, 132, 146, 148
- and manuscript culture, 7, 139, 144, 145, 155
- and manuscript description, 111, 113–14, 159–62, 166–68, 171, 175–176, 179, 183, 187–88, 192–94, 205–6, 216–17
- and manuscript matrix, 123, 137–38
- and manuscript witness, 111, 113–14, 116–117, 118, 124–25, 128–29, 133, 138, 143–44, 147, 153, 203, 209–10
- and metadata, 7, 159–63, 166–68, 170, 173–76, 179–80, 183–90, 195, 192, 203, 206
- and misconceptions, 93
- and monolingualism, 192–95
- and presentation text, 100
- Map of Early Modern London, 48
- Mappae mundi, 50–51, 65
- maps, 1, 5, 45–65, 77, 87
- Martin, Worthy, 7, 174, 205, 207
- Maryland, University of, 226–27
- Material Evidence in Incunabula (MEI), 207
- Matthiesen Gallery, 3
- McCarthy, Sean, 3
- McGann, Jerome, 3, 46
- McGrady, Deborah, 134–35, 138
- mendicants, 5, 70, 74, 76–89
- Metadata Object Description Schema (MODS), 7, 170, 173–78, 180–84, 185–86, 187, 189, 193
- metadata, 3, 51, 52, 65, 93, 108, 140, 157, 159, 160–64, 165n, 167–68, 169–75, 179, 181, 187–88, 190–92, 195, 203–5, 207, 209, 213, 219, 229, 233, 238
- administrative, 7, 161
- descriptive, 7, 64, 159, 162–63, 166, 173, 189, 192–94
- oxygen, 115, 176, 185, 187
- parchment, 114, 165–68, 170, 185
- properties, 207, 210–13, 215–16, 219
- structural, 161–62
- Miller's Tale, 40–41
- Minnis, A. J., 143
- Minor Works of John Lydgate, 6, 94–95, 108–14, 117
- Mirabilia*, 48
- Mississippi, University of, 225–26, 228
- modern English, 123, 133, 137, 138, 141
- Moi, Toril, 123, 133, 137–38, 141
- Mona Lisa, 213
- motet, 6, 124, 126, 141
- mouvance*, 6, 138, 145, 148–50, 179, 186
- Mozart, Wolfgang Amadeus, 212
- National Endowment for the Humanities (NEH), 224, 227–28, 230, 235
- National Gallery of America, 208
- Neatline (software), 46, 52, 57, 62–65
- Netherlands, the, 209
- Neo4j (software), 8, 216–19
- network (concept), 47, 48, 64, 94, 97n, 100, 103–8, 110–11, 113, 117–18, 172, 196–97, 204, 224, 226
- network (digital), 2, 42, 96, 109, 116–18, 164, 173–74, 212, 216, 218, 224
- network archaeology, 219
- New Critics, 137
- New Philology, 1, 145–46, 149
- New York, 159, 197, 209
- Nicholas I (Czar of Russia), 209
- Nichols, Stephen G., 123, 145–46, 149

- Nieuwenhuys, Charles, 209
 Nodegoat (software), 8, 216–19
 nodes, 215–17, 219
 Nonantola, Abbey of, 218
 Norwich Castle Museum and Art Gallery, 8, 229
 Nowvskie, Bethany, 46, 63–64, 94
 Nun's Priest's Tale, 41
- Odoric, 64
 Omeka (software), 46, 51–52, 64
 online platforms, 6, 7, 28, 46–47, 51–52, 62–65, 100, 102–8, 109, 113–14, 119, 125, 127, 138–39, 223, 225–26, 231, 234
 and collaboration, 99, 109
 and development, 6–7, 8, 95–96, 99–100, 102, 107–8, 119, 223, 228, 230, 238
 and social media, 95–96
 ontologies, 93, 203, 205, 210–12, 215, 219
 Ooi, Beth Newman, 234
Orbis Terrarum (T-O) map, 50
 Ore, Christian-Emil, 211–12
 Orosius, 18, 48
 Oxford Friars Project, 78
 Oxford, 69–70, 78–80, 82, 86–87, 89, 165, 184, 237–38
- Palladio, 47
 Paradise of Gathalonabes, 57
 Paris, Gaston, 147, 153
 Paris, 209
 Parker on the Web, 7, 160, 163, 175, 177, 186–93
 Parson's Tale, 40
 Patterson, Lee, 69, 89
 Pearsall, Derek, 150
 Peirce, Charles Sanders, 100–3
 Petrucchi, Giambattista, 218
 Phillips Collection, 204–5, 210, 216–19
 Phillips, Sir Thomas (1792–1872), 204–7, 210, 217–19
 philology. *See also* New Philology, 64, 123, 147, 149, 229
- Pierce the Ploughman's Crede* (Anon.), 5, 69–71, 73, 75–78, 81, 86–87, 89
Piers Plowman (William Langland), 6, 70, 75, 78, 80, 145, 150, 154
 Kane-Donaldson edition, 77, 149–50
Piers Plowman Electronic Archive, 6, 145, 151–54
 Pitti, Daniel, 226–27
 Pittsburgh, 209
 Pliny, 48
 Plumley, Yolanda, 125, 136
 poetry and genre, 4, 11–14, 18–20, 22, 25, 71
 Portable Antiquities Scheme, 229
 portable texts, 125, 129, 139
 power dynamics, 235
 presentation, 1
 digital, 2, 4–5, 93–94, 102, 104–10, 118–19
 textual, 93, 111, 113, 136
 material, 93, 100, 102, 136
 Prester John, 55–57, 59
 Project Andvari, 8, 223–33, 235, 238
 project credit, 8, 224–25, 234–37
 project management, 4, 8, 223–29, 232, 234, 237–39
 Project Paradise, 5, 45–48, 50–51, 62
Prologue (Machaut), 132–33
 property graphs, 205, 215–17, 219
 protégé, 231
 PROV Data Model and Ontology, 203
 provenance, 4, 7–8, 162, 193, 204–13, 215–16, 218–19
Provenance Standard Data Model, 209, 210–13
 Puisse que Desirs me veut dou tout grever (Lo72), 136
 Puisse que Desirs ne me laisse durer (Lo71), 136
 punctualization, 94
 Permanent User Resource Locators (PURLS), 129
 Python (programming language), 14
- Quant en moy / Amour et biauté / Amara valde*, 126

- R (programming language), 4, 18
 radical transparency, 94–95, 106–8
 and design statement, 113–14, 116
 reception, 71, 75, 78, 100, 103, 111, 113, 124
Relatio (by Odoric), 64
 relative frequency, 13–16, 18, 21, 29, 36
 research and critique, 3
 Resource Description and Access (RDA), 207, 221
 Rockwell, Geoffrey, 1
 Roland Park Company Papers, 129–30
 Roma (Rome), 218
Roman de la Rose Digital Library, 123
 Romance language, 218
 Roy Rosenzweig Center for History and New Media, 51
- sales catalogues, 186, 190–91
 Samuels, Lisa, 46
 Saussure, Frederick, 99–101
 scepticism, 5, 78, 237
 Schoenberg Database of Manuscripts, 209–10
 Scholars' Lab, 52
 scribal intention, 153
 semiotics, 94, 99
 and infinite semiosis, 101–3
 and Interpretant, 101–3, 107
 and Object, 100–2, 107
 and Representamen, 101, 107
 and Sign, 99–102
 and *Significant*, 99, 101
 and *Signifie*, 99, 101
 as dyad, 99–100
 as triad, 100–2
 Seth (Biblical figure), 59–62, 65
 Shakespeare, William, 5, 71–75
 shared resources, 224
Siege of Jerusalem Electronic Archive, 6, 145, 152
Siege of Jerusalem, 144, 154
 Simple Event Model (SEM), 215
 Simple Knowledge Organization System (SKOS), 232
 Simpson, James, 75
 Sinclair, Stéfan, 1
Sir Gawain and the Green Knight, 51
Sir Orfeo, 51
 Skeat, Rev. Walter, 70–71
 SketchUp, 82
 Smith, Irwin, 73–76, 86
 Society of American Archivists, 236
 software development
 and “scope creep”, 108
 and code mills, 96
 and development philosophy, 64–65, 78–79, 94
 and solution design document, 108
 and training, 2
 Solinus, 48
 Sotheby's, 216–18
 spatial storytelling, 56–65
Speculum, 145, 149
 stemmatics, 136
 stop words, 15–17, 20, 24
 Stringer, Gary, 125
 Structured Query Language (SQL), 111–12
 Student Collaborators' Bill of Rights, 236
 Stylometrics, 11, 13, 25
 Swedish National Heritage Board, 8, 229
 Swedish Open Cultural Heritage (SOCH), 229
 Symphony No. 41 in C major, K. 551 (by W. A. Mozart), 212
- Taylor, Christopher, 54
 technical narrative, describing, 3
Ten by Twenty, 130
 Text Encoding Initiative (TEI), 6, 114, 118, 124–25, 129, 140, 170–71, 206
 apparatus, 125
 attributes, 125
 elements, 114–115, 125, 128, 130–31, 180, 184–85, 187–88, 206
 parallel segmentation, 127–29
 tags, 114, 130, 169, 185, 187–88
 and P4 Guidelines, 177, 186, 188
 and P5 Guidelines, 125–26, 175, 177–78, 186, 188, 193
 and XML, 114–16, 153, 169–71, 170–72, 174–78, 180–89, 191, 193, 201, 206, 221

- text encoding, 6
- textual analysis, 4, 11
- textual editing, 6
- critical apparatus, 96, 107
 - digital edition, 6, 93
 - editorial theory, 6, 93–94
 - print edition, 6, 65, 70, 93, 114, 124, 128, 138, 139, 140, 144
- The Hague, 209
- The Travels of John Mandeville*. See *The Book of John Mandeville*
- Thorpe, Thomas, 218
- Three Indias, 54
- tools, xii, 1–6, 11, 27–29, 32–34, 36, 42, 69, 82, 96, 119, 223, 233
- as embodiment of ideologies, 6, 78, 93–96, 98, 106
- Trafalgar, the battle of, 212
- transcription, 8, 46, 48, 51, 110–11, 113–16, 119, 146, 153
- translatio*, 143–46
- translation, 5, 49, 52, 125, 127, 145, 190, 193
- between media, 7–8, 124, 130, 136, 141
- transmission, 1, 47, 124, 136–38, 145, 148, 150, 204
- Turville-Petre, Thorlac, 150
- Tuthmosis III, amphora of, 214
- unintended consequences, 96, 99
- Unix, 97
- van Eyck, Jan, 208
- variance, 6–7, 124, 138, 145–51, 154–55
- Virginia, University of (UVA). See [IATH or Scholar's Lab](#)
- virtual facsimile, 4–5, 8, 89, 94, 106, 110
- and adaptation, 94–95, 102
- visualization, 5, 29–30, 35–37, 39, 42, 46–48, 61, 63, 69, 78, 83–88, 113–14, 170, 204, 209–10, 218–19
- Visualizing Medieval Places (VMP), 47–48
- Vitz, Evelyn Birge, 147
- Voyant (software), 4, 28, 36, 38–40, 42
- Walters Art Museum, 7, 160, 175–87, 192–93
- Warton, Thomas, 70
- Washington, D.C., 209, 227–28, 239
- Westrem, Scott, 51, 63
- William II (King of the Netherlands), 209
- William of Boldensele, 64
- William of Tripoli, 64
- Wing, Betsy, 145, 155
- Witmer, Andrew, 3
- WordprocessingML, 115
- World War II, 211
- Wrisley, David J., 47–48
- Wycliffite writing, 70, 74–75, 78–79
- XPath (XML Path Language), 180, 186–88
- XQuery (XML Query), 116, 176, 185–87, 190
- Zumthor, Paul, 6, 145–46, 148–50